CALCULUS OF VARIATIONS

Lectures by

Professor R. Courant

New York University

1945-1946

Contents

INTRODUCTION

The Calculus of   has assumed an increasingly important role in modern developments in analysis, geometry, and physics. Originating as a study of certain maximum and minimum problems not treatable by the methods of elementary calculus, variational calculus in its present form provides powerful methods for the treatment of differential equations, the theory of invariants, existence theorems in geometric function theory, variational principles in mechanics. Also important are the applications to boundary value problems in partial differential equations and in the numerical calculation of many types of problem which can be stated in variational form. No literature representing these diverging viewpoints is to be found among standard texts on calculus of variations, and in this course an attempt will be made to do justice to this variety of problems.

The subject matter with which calculus of variations is concerned is a class of extremum (i.e. maximum or minimum) problems which can be considered an extension of the familiar class of extremum problems dealt with by elementary differential calculus. In the elementary problems one seeks extremal values of a function of one or more (but in any case a finite number) real variables. In the more general problems considered by calculus of variations, the functions to be extremized, sometimes called functionals, have functions as independent variables. The area A(f) below a curve y = f(x), for example, is a functional since its value depends upon a whole function f. (It is possible to treat a functional as a function of an enumerable set of Fourier coefficients, but this attack usually leads to almost insuperable difficulties.)

One of the earliest problems of this type was the isoperimetric problem considered by the ancient Greeks. This is to find, among all closed curves of a given length, the one which encloses the maximum area. It is intuitively evident that the solution is a circle, but this fact has been satisfactorily

proved only in recent times, and the corresponding theorem concerning the sphere is even more difficult.

The modern development of calculus of variations, however, began in 1696 with the formulation of the brachistochrone problem by John Bernoulli. This problem is to find, among all curves connecting two given points, that one which has the property that a particle sliding along it under the action of gravity alone falls from one point to the other in the least time. This problem excited great interest among the mathematicians of that day, and gave rise to a train of research which is still continuing.

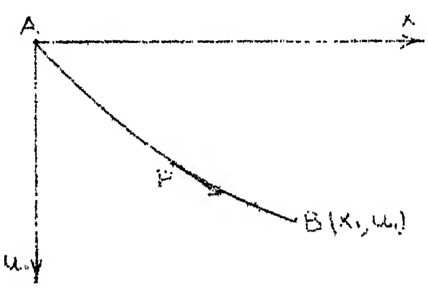
Subsequent developments in classical calculus of variations were the derivation of necessary conditions for an extremum (corresponding to the conditions grad $f(x_1,x_2,\dots x_n)= 0$ for a function f of n variables) by Euler and more rigorously by Lagrange; and the development of sufficient conditions (corresponding to the consideration of the quadratic form in second derivatives of $f(x_1,x_2,\dots x_n)$ at a stationary point) by Hamilton, Jacobi, and others; culminating in the completion of this theory by Weierstrass.

The broader aspects of physical variational principles were first set forth by Maupertuis, and were given a firmer foundation by the work of Euler, Hamilton, Jacobi and Gauss.

We will now consider the mathematical formulation of several problems:

a) The Brachistochrone

A particle P slides under the influence of gravity along a curve connecting A and B. The velocity v at any point is given by



$$v = \frac{ds}{dt} = \sqrt{2gu} ,$$

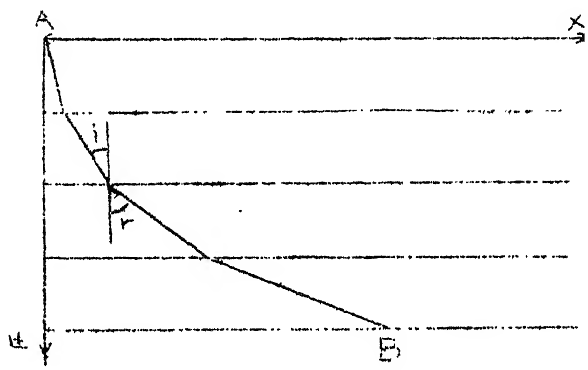
so that the time of fall T is

$$T = \int_{(A)}^{(B)} dt = \int_{(A)}^{(B)} \frac{ds}{\sqrt{2gu}} .$$

Suppose the curve is given by $u = f(x)$, where $f(0) = 0$, $f(x_1) = u_1$, and $f(x)$ is assumed to be piecewise differentiable. Then $ds = \sqrt{1 + u'^2}\,dx$. Hence the solution of the problem can be obtained by finding the function $u = f(x)$ which minimizes the integral T(a functional)

$$T = \frac{1}{\sqrt{2g}} \int_0^{x_1} \sqrt{\frac{1+u'^2}{u}}\ dx.$$

Bernoulli obtained the solution to this problem using an entirely different line of reasoning. He approximated the path $u = f(x)$ by a series of line segments dividing the distance fallen into equal parts, the particle velocity being assumed constant throughout each segment. It is an elementary exercise in calculus to derive Snell's law of refraction



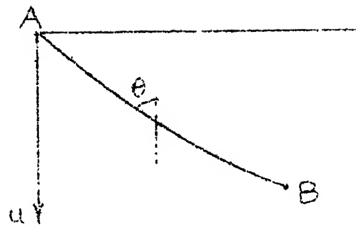
$$\frac{\sin i}{v_i} = \frac{\sin r}{v_r} = \text{const.}$$

as the condition for the path of minimum time across a discontinuity. Taking the limit as the segments are made smaller, Bernoulli argued that the curve would be given by

$$\frac{\sin \theta}{\sqrt{2gu}} = \text{constant}$$

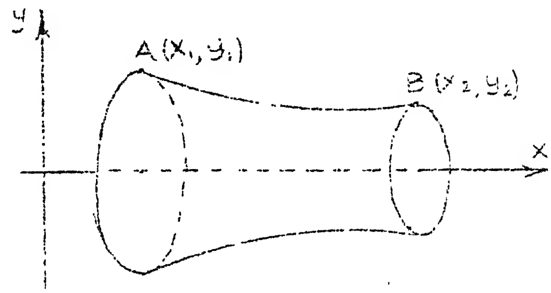
which is indeed the correct answer, characterizing the cycloid. Of course, Bernoulli's solution is only an indication rather

than a proof, since he neither justified the limiting process, nor showed that his solution was unique.



b) Minimum Area of a Surface of Revolution

Consider the surface generated by revolving the curve AB about the x-axis. If the equation of this curve is $y=f(x)$, where $f(x_1)=y$ and $f(x_2)=y_2$, and



f is piecewise differentiable, then the area of the surface is given by the functional

$$I(f) = 2\pi \int_{x_1}^{x_2} f\sqrt{1 + f'^2}\, dx.$$

The problem, then, is to determine f so that I(f) is a minimum. The problem can be "solved" physically by stretching a soap film between the two circles (made of wire) at A and B. Surface tension in the film will minimize the area.

c)

The curve of shortest length connecting two points in a plane is a straight line. This need not be taken as an axiom, but can be proved. Similarly, on the surface of a sphere, the curve of least length is the arc of a great circle. In general, on any surface, the curves of least length connecting pairs of points are called geodesics and their determination leads to problems in calculus of variations. In case the surface is developable (i.e. one which can be deformed into a plane without altering length--e.g. a cone) the geodesics are given by the

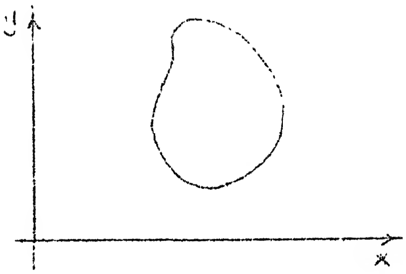
corresponding straight lines in the plane.

d) The Isoperimetric Problem

Consider a plane closed curve given in parametric form by

$x = f(t)$

$y = g(t)$



where f and g have piecewise continuous derivatives, are of period 2π in t, and the curve has a given length L,

$$L = \int_0^{2\pi} \sqrt{\dot{x}^2 + \dot{y}^2}\, dt.$$

The problem is to find, among all f and g satisfying these conditions, the pair that maximizes the area A

$$A = \frac{1}{2}\int_0^{2\pi} (x\dot{y} - y\dot{x})\,dt.$$

This problem is different from the foregoing three problems in that we seek to extremize a functional A of two variables (the functions f and g) subject to a prescribed condition, L = constant. All such problems in calculus of variations are called isoperimetric problems; the analogy with the corresponding elementary problem of extremizing a function F(x,y) of two real variables subject to an auxiliary condition, say G(x,y) = 0, is evident.

We will now prove that the circle, i.e.

$$f(t) = a_1 \sin t + a_2 \cos t$$

$$g(t) = a_2 \sin t - a_1 \cos t$$

maximizes A, subject to L = constant. Consider the expression

$$I = \frac{L^2}{4\pi} - A.$$

For the circle, $I = 0$. We then wish to show that $I > 0$ for all other curves. Let $t = 2\pi \frac{s}{L}$, where s is arc length. Then

$$\dot{x}^2 + \dot{y}^2 = \dot{s}^2 = L^2/4\pi^2,$$

and

$$\frac{L^2}{4\pi} = \frac{1}{2}\int_0^{2\pi} (\dot{x}^2 + \dot{y}^2)\, dt.$$

Hence

$$I = \frac{1}{2}\int_0^{2\pi} \left[(\dot{x}^2 + \dot{y}^2) - (x\dot{y} - y\dot{x})\right] dt$$

$$= \frac{1}{4}\int_0^{2\pi} \left[(\dot{x} + y)^2 + (\dot{y} - x)^2 + (\dot{x}^2 - x^2) + (\dot{y}^2 - y^2)\right] dt.$$

Since $(\dot{x} + x)^2 + (\dot{y} - x)^2 \geq 0$, we will consider

$$I_1 = \int_0^{2\pi} (\dot{x}^2 - x^2)\, dt + \int_0^{2\pi} (\dot{y}^2 - y^2)\, dt.$$

Under the conditions imposed, we may expand x and y in Fourier series

$$x \sim \sum_{n=0}^{\infty} (a_n \cos nt + b_n \sin nt)$$

$$y \sim \sum_{n=0}^{\infty} (a_n' \cos nt + b_n' \sin nt)$$

By taking the center of gravity of the curve as the origin (i.e. translating the axes $x' = x + x_0$, $y' = y + y_0$) so that

$$\int_0^{2\pi} x'\, dt = \int_0^{2\pi} y'\, dt = 0$$

we have

$$a_0 = a_0' = 0.$$

Then, dropping the primes,

$$\int_0^{2\pi} (\dot{x}^2 - x^2)dt = \pi \sum_{n=1}^{\infty} \left[n^2(a_n^2 + b_n^2) - (a_n^2 + b_n^2) \right]$$

which is positive unless $a_n = b_n = 0$ for $n > 1$, i.e. unless

$$x = a_1 \cos t + b_1 \sin t.$$

Similarly
$$\int_0^{2\pi} (\dot{y}^2 - y^2)dt > 0$$

unless

$$y = a_1' \cos t + b_1' \sin t.$$

But in case x and y are both of this form, we have

$$I = \frac{1}{4} \int_0^{2\pi} \left[(\dot{x} + y)^2 + (\dot{y} - x)^2 \right] dt$$

$$= \frac{\pi}{4} \left[(a_1 + a_1' + b_1' - b_1)^2 + (a_1 - a_1' + b_1 + b_1')^2 \right]$$

which is zero only if $a_1 = -b_1'$, $a_1' = b_1$. Hence $I > 0$ unless

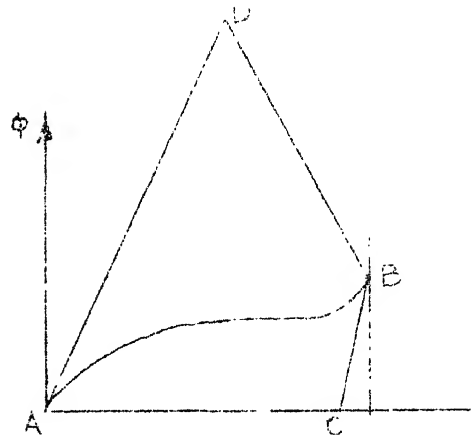
$$x = a_1 \cos t + b_1 \sin t$$

$$y = b_1 \cos t - a_1 \sin t$$

which are the parametric equations of a circle.

In all the problems that we have considered so far we have tacitly assumed that they make sense, i.e., that a solution exists. However, this is by no means always the case. For

example, consider the integral



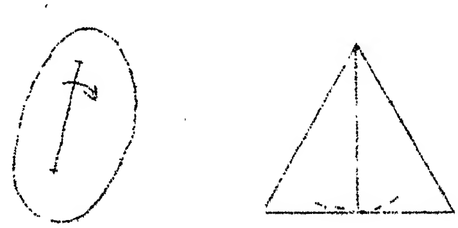
$$\int \frac{dx}{1+[\phi'(x)]^2}$$

where ϕ is subject to the condition that it pass through the two points A and B, and let us try to find a continuous and piecewise differentiable function ϕ which either maximizes or minimizes $I(\phi)$. By inspection we see that

$$0 < I \quad < 1$$

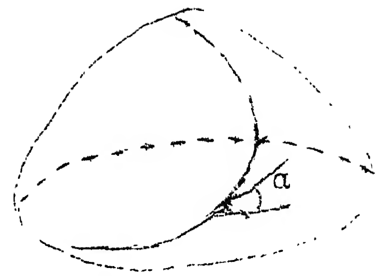
since the integrand is positive and always less than one. However from the figure it is easily seen that by picking point C very close to $x = 1$ we can make I take values as close to unity as we please for the curve A C B, and by taking the ordinate of D large enough we can make I as small as we please for the curve A D B. Since there is no admissible curve ϕ which will make $I(\phi)$ take on the values 0 or 1, there is no solution to either the minimum or maximum problem.

Let us now consider a problem in which the existence of a solution depends on the class of admissible curves. We look for a closed curve of minimum area, within which a line of given length can turn through a complete revolution. If we limit ourselves to convex curves, the solution is given by the equilateral triangle having the given line as altitude. However, if we remove the restriction of convexity, it can be shown that the area can be made as small as we please,



$f(o) = v_o$, $f(\frac{\pi}{2}) = 0$,

$f(\alpha)$ and $f'(\alpha)$ monotonic.



5) Show that any admissible function $\phi(x)$ can be approximated by admissible functions $\phi_1(x)$ and $\phi_2(x)$ such that $I(\phi_1)$ can be made as small as we please and $I(\phi_2)$ as close to unity as we please, where

$$I(\phi) = \int_0^1 \frac{dx}{1 + [\phi'(x)]^2} .$$

FORMALISM OF CALCULUS OF VARIATIONS

1. The Euler Equation. The simplest type of problem in the calculus of variations is to extremize a functional $I(\Phi)$ of only one independent variable, the function $\Phi(x)$. In practice, the functional is usually an integral of a given form, and we will henceforth restrict our discussion to functionals which are integrals. In general terms, then, the simplest type of problem is to extremize

$$(1) \qquad I(\Phi) = \int_a^b F(x, \Phi(x), \Phi'(x))dx \quad .$$

where F is a given function which we will assume has continuous first partial derivatives and piecewise continuous second partial derivatives. The function $\Phi(x)$ will be restricted to the class of admissible functions satisfying the conditions

$$(2) \qquad \begin{aligned} &\Phi(a) = A, \quad \Phi(b) = B \\ &\Phi(x) \text{ continuous} \\ &\Phi'(x) \text{ piecewise continuous} \end{aligned}$$

The brachistochrone is an example of this type of problem.

Assuming that an admissible function $u(x)$ exists for which $I(u)$ is an extremum, we first wish to find a necessary condition which this function must satisfy.

Consider a function $\Phi(x,t)$ such that

$$(3) \qquad \begin{aligned} &\Phi(x,t) \text{ is admissible for all } t \\ &\Phi(x,t) \text{ and } \Phi_t(x,t) \text{ are continuous} \\ &\Phi_{xt}(x,t) \text{ is piecewise continuous} \\ &\Phi(x,0) = u(x) \ . \end{aligned}$$

For example, we may choose $\Phi(x,t) = u(x) + t \ldots$, however, any function satisfying (3) will suffice. If we define

(4) $$G(t) = I(\Phi(x,t)) ,$$

then $G(t)$ has a stationary point at $t = 0$. Accordingly

(5) $$\left.\frac{dG}{dt}\right|_{t=0} = \frac{d}{dt}\int_a^b F(x,\Phi,\Phi')dx \Big|_{t=0} = 0 .$$

Differentiating under the integral sign, we have

(6) $$\int_a^b [F_u\zeta + F_{u'}\zeta']dx = 0 ,$$

where*

(7) $$\zeta(x) = \Phi_t(x,0) .$$

According to (2), (3), and (7) we see that

(8) $$\begin{array}{l} \zeta \text{ is continuous} \\ \zeta' \text{ piecewise continuous} \\ \zeta(a) = \zeta(b) = 0 , \end{array}$$

the last equation being true since $\Phi(a,t) = A$ and $\Phi(b,t) \equiv B$.

If we modify (6) by integrating by parts, the analogy with the corresponding necessary condition for an extremum of a function of n variables is revealed.

$$\int_a^b [F_u\zeta + F_{u'}\zeta']dx = F_{u'}\zeta\Big|_a^b + \int_a^b [F_u - \frac{d}{dx} F_{u'}]\zeta dx$$

(9) $$= \int_a^b [F_u - \frac{d}{dx} F_{u'}]\zeta dx = 0 .$$

In the case of a function $f(x_1,\ldots,x_n)$ of n variables, we may derive the necessary conditions for an extremum by considering $g(t) = f(x_1(t),x_2(t),\ldots x_n(t))$ where the

---

* . In the literature ζ and ζ' are usually called the <u>variation</u> of u and u' respectively, and written δu, $\delta u'$.

equations $x_i = x_i(t)$ define a curve in n-dimensional space. If $(x_1(0), x_2(0), \ldots x_n(0))$ is to be an extremum, then

(10) $$\left.\frac{dg}{dt}\right|_{t=0} = 0 .$$

In other words,

(11) $$\sum_{i=1}^{n} f_{x_i} \cdot \dot{x}_i = 0 .$$

In vector notation, this is the inner product

(12) $$\text{grad } f \cdot \vec{V} = \Theta ,$$

where $\vec{V}$ is the "velocity" vector with components $(\dot{x}_1, \dot{x}_2, \ldots \dot{x}_n)$. This relation must hold for arbitrary V, from which we may conclude that

(13) $$\text{grad } f = \Theta ,$$

which is the desired necessary condition for an extremum. Referring to (9), we see that we have an equation similar to (11), the discrete sum being replaced by an integral, $\dot{x}_i$ and f_{x_i} by ζ and $[F_u - \frac{d}{dx} F_{u'}]$ respectively. By analogy, the necessary condition corresponding to (13) would be

(14) $$F_u - \frac{d}{dx} F_{u'} = 0 .$$

This, indeed, is the well known Euler equation for the extremizing function u. We observe, however, that

(15) $$\frac{d}{dx} F_{u'} = F_{u'u'} u'' + F_{uu'} u' + F_{u'x}$$

does not exist for all admissible functions. Hence the Euler equation (14) does not constitute ~~an~~ a priori satisfactory formulation of a necessary condition since it is not clear in advance that for an <u>extremizing</u> function u(x), the quantity $\frac{d}{dx} F_{u'}$ exists. This difficulty may be avoided by integrating (6) by parts in the following way:

$$\int_a^b [F_u\zeta + F_{u'}\zeta']dx$$

$$(16) \qquad = \zeta\int^x F_u dx \Big|_a^b + \int_a^b \zeta'(F_{u'} - \int^x F_u dx)dx$$

$$= \int_a^b \zeta'(F_{u'} - \int^x F_u dx)dx = 0 .$$

Equation (16) must hold for all ζ such that

$$(17) \qquad \begin{array}{l} \zeta(a) = \zeta(b) = 0 \\ \zeta \text{ continuous} \\ \zeta' \text{ piecewise continuous,} \end{array}$$

For our purposes it will be convient to prove the following <u>Fundamental Lemma</u>: if

$$\int_a^b \zeta'(x)f(x)dx = 0$$

for all ζ satisfying (17), and $f(x)$ is piecewise continuous, then $f(x)$ is a constant.

<u>Proof</u>: Since

$$\int_a^b \zeta' dx = \zeta\Big|_a^b = 0 ,$$

it follows that, for any constant C,

$$(18) \qquad \int_a^b \zeta'(f - C)dx = 0 .$$

In particular, we may choose C so that

$$(19) \qquad \int_a^b (f - C)dx = 0 : \text{ i.e. } C = \int_a^b fdx/(b - a) .$$

With this choice of C, it follows that the function

$$\int_a^x (f - C)dx$$

satisfies (17); hence (18), which must hold for all functions ζ which satisfy (17), must hold in particular for

$$\zeta = \int_a^x (f - C)dx ,$$

i.e. for $\zeta' = f - C$. If we substitute this function in (18) we obtain

$$\int_a^b (f - C)^2 dx = 0 . \tag{20}$$

Since $f(x)$ is piecewise continuous, this implies that $f - C \equiv$ which proves the lemma.

Applying the lemma to (16) we conclude that

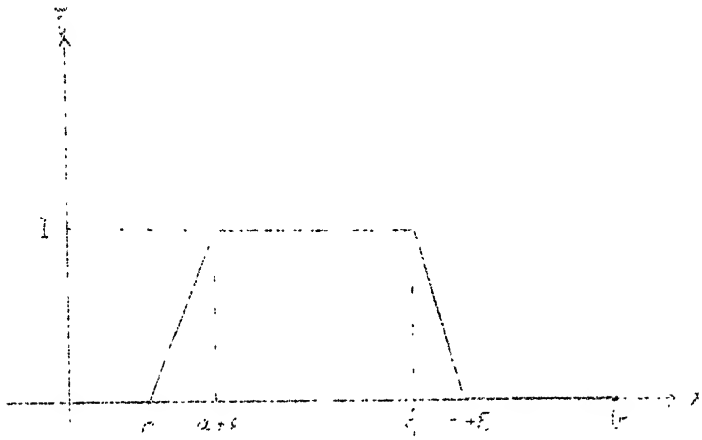
$$F_{u'} - \int^x F_u dx = C . \tag{21}$$

This is the desired necessary condition which an extremizing function u must satisfy. It is, in fact, the first integral of the Euler equation (14). Since F_u is continuous, we may differentiate (21) and conclude that if u is an extremizing function, $\frac{d}{dx} F_{u'}$ exists and is, in fact, continuous, being equal to F_u. Referring to (15) we observe that this implies that u'' is piecewise continuous, provided $F_{u'u'} \neq 0$. In other words, any extremizing function must belong to a subclass of the class of admissible functions -- namely the subclass possessing continuous first and piecewise continuous second derivatives; it follows that the Euler equation (14) is a satisfactory necessary condition for an extremizing function.

It is important, however, to realize that this derivation of (21) is needlessly elaborate. Since all we are seeking is a <u>necessary</u> condition, it follows that any condition which an extremizing function u must satisfy for any particular class of ζ's will be a necessary condition. Now (21) is a relation which must hold for every value of x between a and b, i.e. it is a one parameter set of relations. It therefore seems reasonable that if we choose almost any one parameter family of ζ's, we should be able to derive (21).

For example, let

$$
(22) \qquad \zeta(x) = \begin{cases} \dfrac{1}{\varepsilon}(x - a) & a \leq x \leq a + \varepsilon \\ 1 & a + \varepsilon \leq x \leq \xi \\ 1 - \dfrac{1}{\varepsilon}(x - \xi) & \xi \leq x \leq \xi + \varepsilon \\ 0 & \xi + \varepsilon \leq x \leq b \end{cases}
$$



If we substitute (22) into (6), we obtain

$$
\int_a^{a+\varepsilon} \frac{1}{\varepsilon}(x - a)F_u dx + \int_{a+\varepsilon}^{\xi} F_u dx + \int_{\xi}^{\xi+\varepsilon} [1 - \frac{1}{\varepsilon}(x-\xi)]F_u dx
$$

$$
+ \frac{1}{\varepsilon} \int_a^{a+\varepsilon} F_{u'} dx - \frac{1}{\varepsilon} \int_{\xi}^{\xi+\varepsilon} F_{u'} dx = 0 .
$$

But

$$
\left| \int_a^{a+\varepsilon} \frac{1}{\varepsilon}(x - a)F_u dx \right| \leq \int_a^{a+\varepsilon} |F_u| dx \to 0, \text{ as } \varepsilon \to 0
$$

and similarly for the third term. Also

$$\frac{1}{\varepsilon}\int_a^{a+\varepsilon} F_{u'}dx = F_{u'}(a + \alpha\varepsilon) \ , \qquad (0 \leq \alpha \leq 1)$$

and

$$\frac{1}{\varepsilon}\int_\xi^{\xi+\varepsilon} F_{u'}dx = F_{u'}(\xi + \beta\varepsilon) \ , \qquad (0 \leq \beta \leq 1) \ ,$$

so that letting ε approach zero, we obtain

$$\int_a^\xi F_u dx + F_{u'}\Big|_\xi^a = 0$$

which is equivalent to (21). Many other specific one parameter families can also furnish a derivation of (21).

Problems

1) Give a direct proof that if

$$\int_a^b \zeta(x)f(x) = 0$$

for all ζ satisfying (17) where $f(x)$ is piecewise continuous then $f(x)$ is identically zero.

2) Repeat problem (1) where ζ is restricted to the class of functions having continuous second derivatives.

3) Prove the following generalization of the Fundamental Lemma:

Any piecewise continuous function $f(x)$, for which

$$\int_a^b \zeta^{[k]}(x)f(x)dx = 0$$

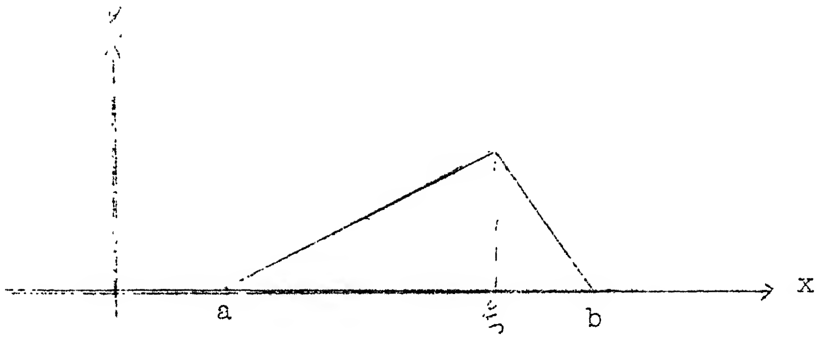
for all $\zeta(x)$ such that ζ, ζ', $\zeta^{[k]}$ are continuous and

$$\zeta^{[n]}(a) = \zeta^{[n]}(b) = 0$$

for $n = 0,1,\ldots k-1$; is a polynomial of degree $k-1$.

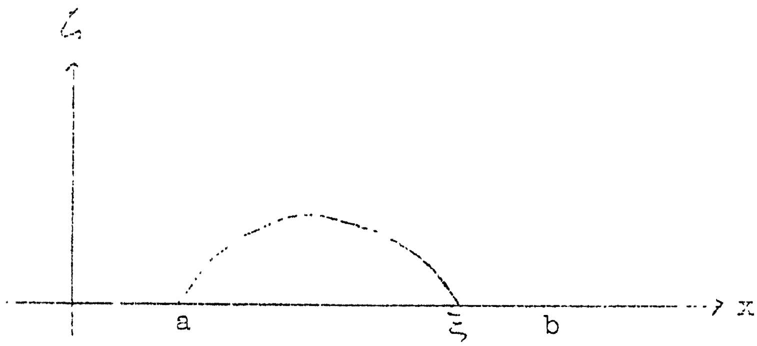
4) Derive Euler's equation using the special variation

$$\zeta(x.\xi) = \begin{cases} (x - a)/(\xi - a) & a \leq x < \xi \\ (b - x)/(b - \xi) & \xi \leq x \leq b \end{cases}$$



5) Derive Euler's equation using the special variation

$$\zeta(x, \xi) = \begin{cases} (x - a)(\xi - x) & a \le x \le \xi \\ 0 & \xi < x \le b \end{cases}$$



6) Show that if $F = F(x,u')$, (i.e. F is not a function of u explicitly), the Euler equation may be solved by quadratures in the form

$$u = \int g(x,C_1)dx + C_2$$

where $u' = g(x,C_1)$ is the explicit solution of the implicit equation

$$F_{u'}(x,u') = C_1 \quad .$$

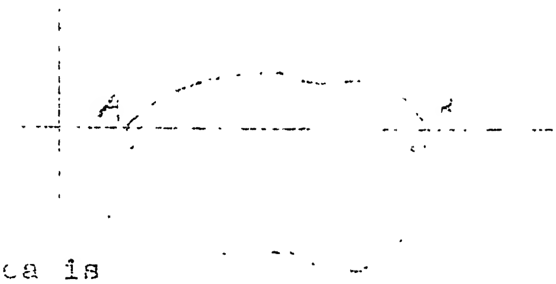
7) Derive a similar solution of the Euler equation for the case $F = F(u,u')$.

8) Use the Euler equation to solve the brachistochrone problem, the problem of minimum surface of revolution, and the problem of the shortest distance between two points of a plane.

9) The isoperimetric problem may be reduced to the Euler equation in the following way. Consider two fixed points A and B on a closed curve. Then, assuming that the closed curve (of given length) encloses maximum area, the area enclosed by the arc $\widehat{AB}$ and the chord AB must certainly be a maximum for all areas of the length of $\widehat{AB}$, say L.

Let the curve AB be given by

$$x = \Phi(s)$$
$$y = \Psi(s)$$
$$s = \text{arc length}$$
$$A:(s=0) \text{ and } B:(s=L) .$$



Then $\dot{\Phi}^2 + \dot{\Psi}^2 = 1$, and the area is

$$A = \int_a^b y\,dx = \int_0^L \Psi \dot{\Phi}\,ds = \int_0^L \Psi \sqrt{1 - \dot{\Psi}^2}\,ds$$

where $\Psi(0) = \Psi(L) = 0$.

To extremize $A(\Psi)$ is now a simple exercise.

Obtain Ψ, and Φ by use of the Euler equation, and then show that if b is allowed to vary, but L is kept constant, the maximum of A will be for a semi-circle.

2. Generalizations of the Euler Equation. The Euler equation (14) may be generalized along three lines by altering the form of F to contain

a) more than one dependent variable

b) more than one independent variable

c) higher than first derivatives

any combination of

a) Suppose F contains two dependent variables, so that

(23) $$I = I(\Phi,\psi) = \int_a^b F(x, \Phi, \Phi', \psi, \psi')dx ,$$

where Φ and ψ are admissible functions. In order that $I(u,v)$ be an extremum, it is necessary that $I(u,v)$ be an extremum with respect to u and v considered independently. In other words, a necessary condition is the pair of Fuler equations

(24) $$F_u - \frac{d}{dx} F_{u'} = 0$$
$$F_v - \frac{d}{dx} F_{v'} = 0 \quad .$$

In general, if F contains n dependent variables there will be n Euler equations.

An example of such a system is found in the Hamilton Principle of Least Action. Briefly, this says that if a mechanical system be described by n independent coordinates $q_1, \ldots q_n$ (n degrees of freedom), then the motion of the system -- i.e. the determination of $q_i = q_i(t)$ -- will be such as to minimize the "action" I,

(25) $$I = \int_{t_o}^{t_1} [T(q_1, \ldots q_n, \dot{q}_1, \ldots \dot{q}_n) - U(q_1, \ldots q_n)]dt \quad ,$$

where T and U are given functions, the kinetic and potential energy respectively. (Actually, the integral is not an extremum, but only stationary.) The Euler equations which must be satisfied to make I stationary are

(26) $$\frac{\partial(T - U)}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_i}\right) = 0 \qquad i = 1, 2, \ldots n \ .$$

These are, in fact, theLagrangian equations of motion. (For one degree of freedom, this reduces to familiar Newton equation $F = ma$.)

b) Next let us suppose that F contains two independent variables x, y so that

(27) $$I(\Phi) = \iint_R F(x, y, \Phi, \Phi_x, \Phi_y)dxdy ,$$

where R is a given closed connected region in the x,y plane, bounded by the curve C. The function will be restricted to the class of <u>admissible functions</u>

(28)
Φ continuous in R
Φ_x, Φ_y piecewise continuous in R
Φ takes given values on C

Suppose that there is an admissible function u(x,y) for which I(u) is an extremum. If we write

(29) $$\Phi(x,y,t) = u(x,y) + t\zeta(x,y)$$

where ζ is admissible except that $\zeta = 0$ on C, then Φ is admissible for all values of t, and

(30) $$G(t) = I(\Phi(x,y,t))$$

is an extremum at t = 0. Accordingly,

(31) $$\left.\frac{dG}{dt}\right|_{t=0} = \iint_R [\zeta F_u + \zeta_x F_{u_x} + \zeta_y F_{u_y}]dxdy = 0 .$$

Since $\zeta = 0$ on C, this reduces upon integration by parts to

(32)* $$\iint_R \zeta [F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y}]dxdy = 0 ,$$

for all admissible ζ. We conclude that the Euler equation in this case is

(33) $$F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} = 0 .$$

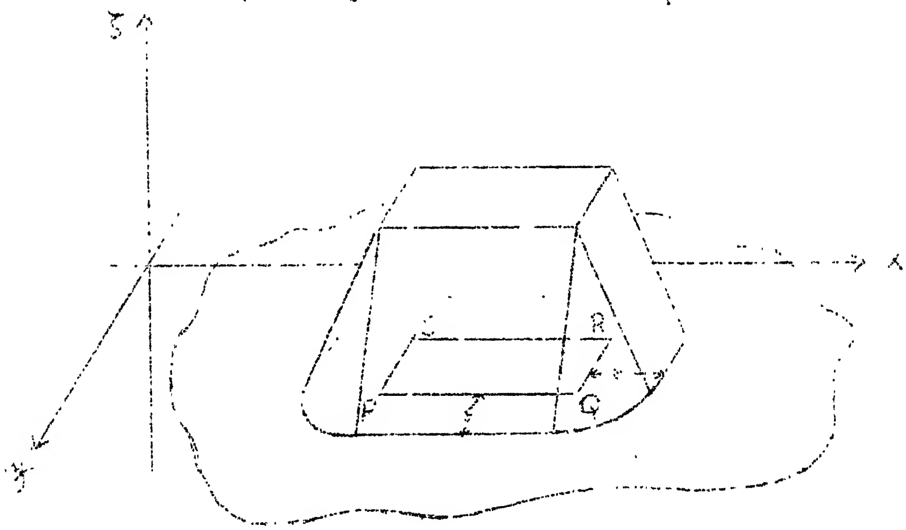
We observe, however, that, as for the simple Euler equation (14) of the preceding section, the derivation of (33) implies the existence of <u>second</u> partial derivatives of u, and hence is inapplicable to some of the admissible functions. In the

---

* $\partial G/\partial x$ is taken to mean the derivative with respect to x, holding y, but not u or its derivatives constant:
$$\frac{\partial G}{\partial x} = G_x + G_u u_x + G_{u_y} u_{xy} + G_{u_x} u_{xx}.$$

earlier case we were able to integrate by parts in a different order and show that for an extremizing function u'' must exist. We are unable to do the corresponding thing here and hence (35) is a necessary condition only if it can be shown that an extremizing function possesses second partial derivatives. We may, however, derive an equation corresponding to the first integral (21) of the Euler equation (14). To do this we investigate (31) directly by considering the following special variation $\zeta(x,y)$; let $\overline{PQRS}$ be an arbitrary rectangle with sides parallel to the coordinate axes and contained inside R, and let $\mu(x,y)$ be the distance from (x,y) to $\overline{PQRS}$; then we define

$$(36)\qquad \zeta(x,y) = \begin{cases} 1 & \text{if } x,y \text{ in } \overline{PQRS} \\ 1 - \mu/\varepsilon & \text{if } 0 \leq \mu \leq \varepsilon \\ 0 & \text{if } \mu > \varepsilon \end{cases}$$



Then ζ is admissible for any $\varepsilon > 0$, and

$$\begin{aligned} \zeta_x &= 0 \text{ on } \overline{PQ},\ \overline{RS} \\ \zeta_y &= 0 \text{ on } \overline{PS},\ \overline{RQ} \\ \zeta_x &= \varepsilon \text{ on } \overline{PS},\ = -\varepsilon \text{ on } \overline{QR} \\ \zeta_y &= \varepsilon \text{ on } \overline{RS},\ = -\varepsilon \text{ on } \overline{PQ} \end{aligned}$$

If we substitute this function ζ into (31), use the theorem of the mean, and let $\varepsilon \to 0$ we obtain

$$\iint_{\square} F_u \, dxdy + \int_P^S F_{u_x} dy - \int_Q^R F_{u_x} dy + \int_P^R F_{u_y} dx - \int_S^R F_{u_y} dx = 0$$

This may be conveniently written

$$(37) \qquad \iint_{\square} F_u \, dxdy = \oint_{\square} (F_{u_x} dy - F_{u_y} dx) .$$

Equation (37) corresponds in our present case to the first integral (21) of the Euler equation. It no longer contains an arbitrary function ζ, but now contains an arbitrary rectangle. We may approximate any region by rectangles and hence conclude that for any rectifiable closed curve C' bounding a region R' in R and for any extremizing u,

$$(38) \qquad \iint_{R'} F_u \, dxdy = \oint_{C'} (F_{u_x} dy - F_{u_y} dx) .$$

This is known as <u>Haar's lemma</u>.

The chief value of this lemma lies in the fact that it is applicable to any admissible function since it does not presuppose (as does the Euler equation (33)) the existence of second partial derivatives. If we know that the extremizing function u has continuous second partial derivatives, then the Euler equation (33) will follow directly from (38) by use of the Green's formula

$$(39) \qquad \oint_{C'} P dx + Q dy = \iint_{R'} (Q_x - P_y) dxdy$$

In case F contains n independent variables $x_1, \ldots x_n$, the equation corresponding to (33) is

$$(40) \qquad F_u - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} F_{u_{x_i}} = 0$$

where here, as before, the condition is necessary only if the extremizing function possesses second partial derivative

Two important examples of this type of problem are the Dirichlet and Plateau problems, namely to minimize

$$(41)\qquad \text{a)}\ \iint_R (u_x^2 + u_y^2)\,dxdy$$

$$\text{b)}\ \iint_R \sqrt{1 + u_x^2 + u_y^2}\,dxdy\ ;$$

the first integral representing, say, the potential energy of an electrostatic field, and the second the area of a surface projecting on R. With given boundary values on C, the latter problem becomes that of minimizing the area of a surface having a given closed space curve as its boundary. The Euler equations corresponding to (41) are

$$(42)\qquad \text{a)}\ u_{xx} + u_{yy} = 0$$

$$\text{b)}\ (1 + u_y^2)u_{xx} - 2u_x u_y u_{xy} + (1 + u_x^2)u_{yy} = 0$$

The first equation is the well known Laplace equation, whose solutions are harmonic functions. If we treat the Dirichlet problem from the point of view of Haar's lemma, we get the condition

$$(43)\qquad \oint_{C'} (u_x dy - u_y dx) = 0$$

for an extremizing function. This must hold for all closed rectifiable C''s in R and hence $u_x dy - u_y dx$ must be an exact differential. It follows that another function v exists such that

$$(44)\qquad u_x = v_y\ ;\quad u_y = -v_x\ .$$

These are the Cauchy Riemann equations which a harmonic function must satisfy. (In particular, assuming that second derivatives of u and v exist, the Laplace equation (42a) follows from (44) by elimination of v.)*

c) Finally, let us suppose that F contains a second derivative so that

$$(45)\qquad I(\Phi) = \int_a^b F(x, \Phi, \Phi', \Phi'')\,dx$$

* Actually, if (44) holds, continuity of the first derivatives alone will insure (42a)--as well as existence of all higher derivatives.

the class of admissible functions ξ now being such that

$$\begin{array}{ll} (46) & \begin{array}{l} \xi,\ \xi' \text{ continuous} \\ \xi'' \text{ piecewise continuous} \\ \xi(a) = A_1, \quad \xi(b) = B_1 \\ \xi'(a) = A_2, \quad \xi'(b) = B_2. \end{array} \end{array}$$

We again suppose, that for the function u, $I(u)$ is an extremum and set $\xi = u + t\zeta$, where ζ is admissible but $\zeta = \zeta' = 0$ at a and b. Then, if

$$G(t) = I(u + t\zeta) ,$$

it follows that $G(0)$ is an extremum and so

$$(47) \qquad \left.\frac{dG}{dt}\right|_{t=0} = \int_a^b [\zeta F_u + \zeta' F_{u'} + \zeta'' F_{u''}]dx = 0 .$$

Integrating by parts, we have

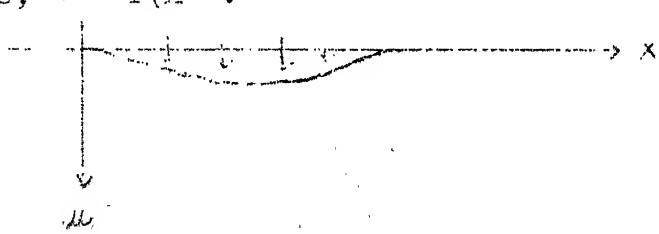
$$(48) \qquad \int_a^b \zeta[F_u - \frac{d}{dx} F_{u'} + \frac{d^2}{dx^2} F_{u''}]dx = 0 .$$

Hence the Euler equation in this case is

$$(49) \qquad F_u - \frac{d}{dx} F_{u'} + \frac{d^2}{dx^2} F_{u''} = 0 .$$

Expanding $(d^2/dx^2)F_{u''}$ we observe that (49) is an equation of the fourth order, and implies the existence of u''''. As in the first section (page 13) it can be shown, for example by considering a special variation ζ, that <u>for an extremizing function</u> u'''' must exist and hence that (49) must hold.

As an example of equation (49) we consider the problem of determining the deflection of a loaded elastic beam. Suppose the beam is clamped rigidly at 0 and 1, and is loaded with a weight per unit length which varies along the x axis, $W = f(x)$.



Then if $u(x)$ is the deflection of the beam at a point x, the potential energy of the system is given by

$$\text{(50)} \qquad \text{P.E.} = \int_0^1 [\tfrac{1}{2}\, \alpha u''^2 - uf(x)]dx \,,$$

where α is a constant determined by the physical properties of the beam. Since the beam is rigidly clamped at the ends, we have $u = u' = 0$ at 0 and 1.

It follows from a basic principle of mechanics that the equilibrium deflection of the beam will be such as to minimize the potential energy. Hence, if u is the deflection, it must satisfy the Euler equation for minimizing (50), namely

$$\text{(51)} \qquad \alpha u'''' - f(x) = 0 \,.$$

This is a fourth order equation whose integration introduces four constants. These may be determined from the four end conditions $u = u' = 0$ at 0 and 1.

In general, if F contains the n'th derivative $\Phi^{[n]}(x)$, so that

$$\text{(52)} \qquad I(\Phi) = \int_a^b F(x, \Phi, \Phi', \ldots \Phi^{[n]})dx \,,$$

then an extremizing function u will satisfy the Euler equation

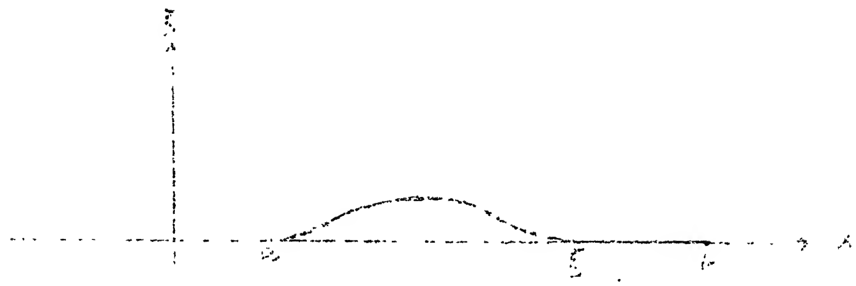
$$\text{(53)} \qquad F_u - \frac{d}{dx} F_{u'} + \frac{d^2}{dx^2} F_{u''} + \ldots (-1)^n \frac{d^n}{dx^n} F_{u^{[n]}} = 0 \,.$$

<u>Problems</u>:

1) Derive the Euler equation (49) assuming that only u'' is piecewise continuous, using integration by parts and the result of problem 3, page 17 .

2) Do problem (1) using the special variation

$$\zeta(x) = \begin{cases} (x - a)^2(x - \xi)^2 & a \le x \le \xi \\ 0 & \xi < x \le b \end{cases}$$



3) Derive the Euler equation (33) from Haar's lemma, assuming that u possesses all second partial derivati

3. Natural Boundary Conditions. In section 1, we considered the stationary values of

$$I(\Phi) = \int_a^b F(x,\Phi,\Phi')dx$$

where Φ was required to be a continuous function with piecewise continuous derivative and such that

$$\Phi(a) = A, \quad \Phi(b) = B .$$

Suppose we now drop this last condition leaving the values of $\Phi(a)$ and $\Phi(b)$ open. Considering

$$G(t) = I(u + t\zeta)$$

in the usual way; we conclude that

$$\left.\frac{dG}{dt}\right|_{t=0} = \int_a^b [F_u\zeta + F_{u'}\zeta']dx = 0 .$$

However, we may no longer require that $\zeta(a) = \zeta(b) = 0$. Hence, the integration by parts gives, for every ζ,

$$F_{u'}\zeta\Big|_a^b + \int_a^b [F_u - \frac{d}{dx} F_{u'}]\zeta dx = 0 .$$

Consideration of a family of ζ's such that $\zeta(a)$ $\zeta(b) = 0$ yields the Euler equation as before. There remains

$$F_{u'}\zeta\Big|_a^b = 0 ,$$

and since $\zeta(a)$ and $\zeta(b)$ are arbitrary, we conclude that

(34) $$F_{u'}\Big|_{x=a} = F_{u'}\Big|_{x=b} = 0 .$$

Thus, if we do not prescribe any end point values for the extremizing function u, we find that such a function must automatically satisfy a relation at each end anyway--this is the so-called natural boundary condition. It is evident a priori, that the Euler equation is a necessary condition whether or not boundary conditions are imposed, since any extremum, if it exists, will have definite boundary values.

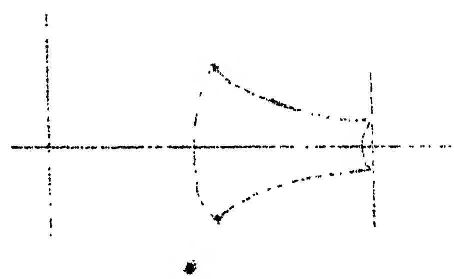
For example, in the problem of the minimum area of revolution, $F = u\sqrt{1 + u'^2}$ and

$$F_{u'} = \frac{uu'}{\sqrt{1 + u'^2}} .$$

If we prescribe $u(a) = A$, i.e. fix one end, but leave $u(b)$ free, we obtain the natural boundary condition

$$u(b)\cdot u'(b) = 0 .$$

Hence the curve which gives the minimum area of revolution will (if b-a is small) be that one which satisfies the Euler equation and is horizontal at $x = b$. If we think of the soap film analogy of this problem, we see that the natural boundary condition is that the film be perpendicular to a wall at b.



In case F contains higher derivatives the same phenomenon occurs. Instead of deriving general conditions, we will consider the specific example of determining the deflection of a loaded beam which is clamped at one end and at the other end is respectively clamped, free, or supported. Referring to the discussion on page 25, we know that the

equilibrium deflection u(x) of the beam will be such as to minimize

$$\text{P.E.} = \int_0^1 \left[\tfrac{1}{2}\, a(u'')^2 - u\cdot f(x)\right] dx. \tag{55}$$

Here the prescribed end conditions are $u = u' = 0$ at $x = 0$, and

i) $u = u' = 0$ at $x = 1$ (clamped)

ii) no condition at $x = 1$ (free)

iii) $u = 0$ at $x = 1$ (supported)

Case i) has already been discussed (see page 26). It leads to the Euler equation of fourth order

$$au'''' - f(x) = 0 \tag{56}$$

the four end conditions necessary to determine a specific solution being prescribed.

We consider case ii) directly, i.e. let $\Phi = u + t\chi$, where $\chi = \chi' = 0$ at $x = 0$, but χ and χ' are left free at $x = 1$. Then if u minimizes (55) we will have

$$\frac{d}{dt}\Big|_{t=0} \int_0^1 [\tfrac{\alpha}{2}\Phi''^2 - \Phi f(x)]dx = \int_0^1 (\alpha u''\zeta'' - f\zeta)dx = 0 .$$

Integrating by parts, and making use of the condition $\zeta(0) = \zeta'(0) = 0$, we have

$$\alpha[u''\zeta'\Big|_{x=1} - u'''\zeta\Big|_{x=1}] + \int_0^1 [\alpha u'''' - f]\zeta dx = 0 .$$

Hence, the Euler equation (56) must still hold. But in addition we must have, for all ζ,

$$u''\zeta' = u'''\zeta = 0 \quad \text{at } x = 1 .$$

In other words, if we prescribe no end conditions at $x = 1$--leave it free--we are automatically led to the "natural" boundary conditions

(57) $$u'' = u''' = 0 \qquad \text{at } x = 1$$

We still have the necessary four boundary conditions to determine the specific integral of (56).

In case iii), where we prescribe $u(1) = 0$ but leave $u'(1)$ free, we must prescribe $\zeta(1) = 0$, but leave $\zeta'(1)$ free. Then it follows that

$$u'' = 0 \quad \text{at } x = 1 ,$$

but no longer that $u''' = 0$ there. Again have four boundary conditions--three prescribed and one "natural"

A second type of free end condition is to leave the coordinates a and b themsleves free, prescribing say only that Φ have end points on a given curve. Suppose, for simplicity, that we fix one end point a, $\Phi(a) = A$, but require only that $\Phi(b) = g(b)$ is a given function.

Supposing that $u(x)$ is an extremizing function which intersects $g(x)$ at $x = b(0)$, we let $\Phi = u + t\zeta$ and $b = b(t)$. Then if

$$G(t) = \int_a^{b(t)} F(x,\xi,\xi')dx \;,$$

we have

$$(58)\qquad \frac{dG}{dt}\Big|_{t=0} = \int_a^{b(0)} [F_u\zeta + F_{u'}\zeta']dx + F(x,u,u')\Big|_{x=b(0)} \cdot \dot{b}(0) = 0$$

Since

$$(59)\qquad u(b(t)) + t\zeta(b(t)) = g(b(t)) \;,$$

differentiating and setting $t = 0$, we conclude that

$$(60)\qquad \dot{b}(0) = \frac{\zeta(b(0))}{g'(b(0)) - u'(b(0))} \;.$$

Substituting this in (58) and integrating by parts we obtain, for all ζ

$$(61)\qquad \int_a^{b(0)} [F_u - \frac{d}{dx}F_{u'}]\zeta dx + [F_{u'} + \frac{F}{g' - u'}]\zeta\Big|_{x=b(0)} = 0 \;.$$

Hence an extremizing function u must satisfy the usual Euler equation, a fixed end condition $u(a) = A$, plus a condition

$$(62)\qquad F_{u'} + \frac{F}{g' - u'} = 0 \quad \text{at } x = b \;.$$

Condition (62) is called the <u>Transversality Condition</u>. Together with $u(b) = g(b)$, $u(a) = A$, and the Euler equation, it determines the point b and the solution u. As before leaving the end condition free results in an automatic end condition--in this case a relation between u and u'. The transversality condition (62) reduces to the previously derived natural boundary condition (54) for the case of b fixed, i.e. the fixed curve is $x = b$; and $g'(b) = \infty$, so that the second term in (62) drops out leaving only

$$F_{u'}\Big|_{x=b} = 0 \;.$$

In many of the specific examples considered so far, F was of the

$$F = P(x,u)\sqrt{1 + u'^2} \; .$$

In this case the transversality condition (62) reduces to

$$u'g' = -1 \; ,$$

in other words the extremizing curve u must be orthogonal to the given curve g.

Problems

1) Show that the natural boundary condition at a free boundary for $F = F(x,\phi,\phi',\phi'')$ is

$$u' - \frac{d}{dx} F_{u''} = 0$$
$$F_{u''} = 0 \; .$$

2) The condition that ϕ have its end point, $x = b$, on a fixed curve g may be reduced to the free end condition for fixed b by transforming the x, y plane so that g becomes a vertical line. Derive the transversality condition in this way.

4. Degenerate Euler Equation. The Euler equation for the simplest problem is

(14) $$F_u - \frac{d}{dx} F_{u'} = 0 \; .$$

In case F is linear in u', i.e. of the form

(63) $$F(x,u,u') = A(x,u)u' + B(x,u) \; ,$$

the Euler equation reduces to

(64) $$A_x - B_u = 0 \; .$$

This is no longer a differential equation but is, in fact, an implicit relation which in general will define u as a function of x. It follows that u(a) and u(b) may not b arbitrarily prescribed in general.

The converse is also true, i.e. if the Euler equation degenerates from a differential equation into an ordinary equation then F must be of the form indicated in (63). For, expanding (14) we have

(65) $$F_u - F_{u'x} - F_{u'u} \cdot u' - F_{u'u'} \cdot u'' = 0 \quad .$$

If (65) is not a differential equation, then the term containing u'' (there is only one) must disappear, i.e.

(66) $$F_{u'u'} = 0 \quad .$$

But (66) implies that F is linear in u', i.e. is of the form indicated in (63). Hence, a necessary and sufficient condition that the Euler equation degenerate is that F be linear in u'.

An important special case of this degeneration is when (63) is satisfied, but also $A(x,u)$ and $B(x,u)$ satisfy (64) identically. Then <u>any</u> u will satisfy the degenerate Euler equation and hence will extremize $I(u)$. In this case

$$I(u) = \int_a^b A du + B dx \quad ,$$

is a line integral whose value is independent of the path of integration (i.e. of u) since we are supposing that (64) is satisfied identically. In other words, $I(u)$ is a constan and so has no proper extremum.

In case F contains a second derivative, the Euler equation is of fourth order

(67) $$u''''F_{u''u''} + u''' [\ldots] + \ldots = 0 \quad .$$

Hence it will degenerate to a lower order if

$$F_{u''u''} = 0 \quad .$$

i.e. if F is linear in u''

(68) $$F(x,u,u',u'') = A(x,u,u')u'' + B(x,u,u') \quad .$$

In two dimensions following the same reasoning, we conclude that the Euler equation (33) will degenerate if and only if F is linear in u_x and u_y,

(69) $$F(x,y,u,u_x,u_y) = A(x,y,u) + B(x,y,u)u_x + C(x,y,u)u_y$$

If (69) holds, the Euler equation becomes

(70) $$A_u - B_x - C_y = 0 ,$$

which in general defines u implicitly as a function of x and y.

Problems

1) Show that (67) is of at most second order if (63) holds.
2) Under what circumstances will (67) degenerate even more (i.e. be of less than second order)?
3) Prove that if and only if F is of the form*

$$F = \frac{\partial L}{\partial x}(x,y,u) + \frac{\partial M}{\partial y}(x,y,u) ,$$

then the degenerate Euler equation (70) becomes an identity. As before this means that I(u) is a constant.
4) Show that the Euler equation for two independent variables and second derivatives will degenerate if and only if F is of the form

$$F(x,y,u,u_x,u_y,u_{xx},u_{xy},u_{yy}) =$$
$$P(x,y,u,u_x,u_y)[u_{xx}u_{yy} - u_{xy}^2] + Q(x,y,u,u_x,u_y) .$$

---

* See footnote, page 21.

5. Isoperimetric Problems. Isoperimetric problems in calculus of variations are concerned with extremizing integrals subject to some sort of auxiliary conditions on the dependent functions Φ, Ψ etc. These conditions may be given in the form of integrals, functional relationships, or differential equations:

a) $\int_a^b G(x,\Phi, \Phi', \Psi, \Psi', \dots)dx = 0$

b) $G(x,\Phi, \Psi, \dots) = 0$

c) $G(x,\Phi,\Phi', \Psi, \Psi', \dots) = 0$.

For the corresponding problems in elementary calculus, a necessary condition for an extremum is given by the Euler-Lagrange rule (sometimes called the method of Lagrangian multipliers). We recall that according to this rule, if, among all points (x,y) such that $g(x,y) = 0$, the point (x_0,y_0) extremizes $f(x,y)$, then--providing $g_x(x_0,y_0)$ and $g_y(x_0,y_0)$ do not both vanish--it follows that (x_0,y_0) is found among the stationary points of $f(x,y) + \lambda g(x,y)$ considered as a function of two independent variables, where the constant λ is determined such that $g(x_0,y_0) = 0$. In other words x_0, y_0, and λ must satisfy the three equations

$$(72)\qquad \begin{cases} f_x(x_0,y_0) + \lambda g_x(x_0,y_0) = 0 \\ f_y(x_0,y_0) + \lambda g_y(x_0,y_0) = 0 \\ g(x_0,y_0) = 0 . \end{cases}$$

We shall generalize this rule to cover the present problems in calculus of variations.

a) We first consider the simplest problem, with an integral side condition. Stated explicitly, we are required to extremize

$$(73) \qquad I(\Phi) = \int_a^b F(x,\Phi,\Phi')dx \quad .$$

among all admissible Φ (see page 8), which satisfy the additional restriction that

$$(74) \qquad K(\Phi) = \int_a^b G(x,\Phi,\Phi')dx = K_o \text{ (constant)} .$$

We suppose that there is such a function u, extremizing (73) subject to (74). Let $\Phi = u(x) + t_1\zeta_1(x) + t_2\zeta_2(x)$ where ζ_1 and ζ_2 are admissible but vanish at a and b. Then

$$(75) \qquad \begin{aligned} I &= I(t_1,t_2) \\ K &= K(t_1,t_2) . \end{aligned}$$

Since we require that Φ always satisfy (74), the parameters t_1 and t_2 are not independent, but must satisfy $K(t_1,t_2) = K_o$. From our hypothesis that u is an extremizing function it follows that I(0,0) is an extremum of $I(t_1,t_2)$ for all t_1,t_2 which satisfy $K(t_1,t_2) = K_o$. Hence we may apply the Euler-Lagrange rule for ordinary functions to conclude that -- providing K_{t_1} and K_{t_2} do not both vanish at (0,0)--there must be a number λ so that

$$(76) \qquad \frac{\partial}{\partial t_i}[I + \lambda K] = \int_a^b \zeta_i[F + \lambda G]_u dx = 0$$

$$i = 1,2.$$

where we define the Euler operator

$$(77) \qquad [H]_u = H_u - \frac{d}{dx} H_{u'} .$$

Since (76) must hold for all ζ_1 and ζ_2, we conclude that

$$(78) \qquad [F + \lambda G]_u = 0 .$$

The restriction that K_{t_1} and K_{t_2} do not both vanish at (0,0) means that

$$\dot{K}_i\Big|_0 = \int_a^b Z_i[G]_u dx \neq 0 \,, \qquad i = 1 \text{ or } 2 \,.$$

or

(79) $$[G]_u \not\equiv 0 \,.$$

The required generalization of the Euler-Lagrange rule is hence the following: if $I(u)$ is an extremum subject to $K(u) = K_o$, and $[G]_u \neq 0$, then there is a constant λ such that $[F + \lambda G]_u = 0$. In general (78) is a second order differential equation containing a parameter λ. Its solution will be in the form

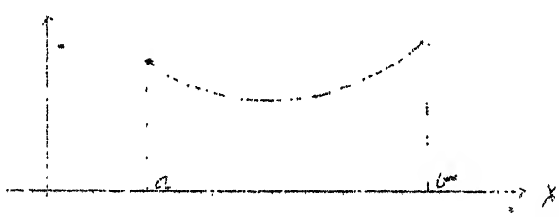
$$u = u(x,\lambda)$$

since the two integration constants may be determined from the (fixed or natural) boundary conditions. The value of λ is then determined so that

$$K(\lambda) = \int_a^b G(x,u(x,\lambda),u'(x,\lambda))dx = K_o.$$

As an illustration, consider the problem of determining the shape of a chain of given length L hanging under gravity.

If $u(x)$ is the shape taken by the chain,



then u is such as to minimize the potential energy

$$PE = \int_a^b u ds = \int_a^b u\sqrt{1 + u'^2}\, dx \,,$$

subject to the usual admissibility conditions, plus the additional restriction that the length

$$\int_a^b \sqrt{1 + u'^2}\, dx = L$$

is a given constant. Observe that the integral to be minimized is the same as the one for the problem of minimum area of revolution, for which the solution was seen to be a catenary. In this problem $F = u\sqrt{1 + u'^2}$, $G = \sqrt{1 + u'^2}$, so the generalized Euler-Lagrange rule states that there is a constant λ such that

$$(80) \qquad [F + \lambda G]_u = \sqrt{1 + u'^2} - \frac{d}{dx} \cdot \frac{u'(u + \lambda)}{\sqrt{1 + u'^2}} = 0 .$$

Solving (80), with $u(a) = u_0$, $u(b) = u_1$, we would get $u = u(x,\lambda)$. Then λ would be determined so that

$$\int_a^b \sqrt{1 + u'^2(x,\lambda)}\, dx = L ,$$

and the result would be a catenary of length L between the points a and b. The side condition is therefore seen to be merely a restriction on the class of admissible functions.

In case F contains more dependent variables, or there are more side conditions we have: if $u_1, \ldots u_k$ extremize $I(u_1, \ldots u_k)$ subject to the n conditions

$$K_i = \int_a^b G_i(x, u_1, u_1', \ldots u_k, u_k')dx = 0 , \qquad i = 1,2,\ldots n ,$$

then, in general, there are n constants $\lambda_1, \ldots \lambda_n$ such that

$$[F + \lambda_1 G_1 + \ldots \lambda_n G_n]_{u_j} = 0 \qquad j = 1,2,\ldots k .$$

b) We next consider the problem of extremising

$$(81) \qquad I(\Phi, \Psi) = \int_a^b F(x, \Phi, \Phi', \Psi, \Psi')dx$$

where Φ and Ψ are subject to the side condition

$$(82) \qquad G(x,\Phi,\Psi) = 0 \ .$$

It will be shown that the generalized Euler-Lagrange rule still applies here, except that λ is no longer a constant, but instead is, in general, a function of x. A proof of this may be obtained directly by eliminating, say, Ψ from (82)--i.e. obtain $\Psi = \Psi(x,\Phi)$--and so reducing the problem to that of the simplest case with no side conditions. We prefer to consider the variation of I.

Suppose, then, that $I(u,v)$ is an extremum under the restriction that $G(x,u,v) = 0$. Let $\Phi(x,t) = u(x) + t\zeta_1(x)$ where $\zeta_1(a) = \zeta_1(b) = 0$, and $\Psi(x,t) = v(x) + \zeta_2(x,t)$ where ζ_2 is determined, once ζ_1 is given, by the relation

$$G(x,u + t\zeta_1,v + t\zeta_2) = 0$$

automatically insuring that $\zeta_2(x,0) = 0$. We know that $I(\Phi,\Psi)$ is an extremum for $t = 0$, and $G(x,\Phi,\Psi) \equiv 0$ (in t) so that for any function $\lambda(x)$

$$\frac{d}{dt}\int_a^b [F + \lambda G]dx\Big|_{t=0} = 0 \ .$$

Accordingly, we have

$$(83) \quad \int_a^b \left\{\zeta_1[F + \lambda G]_u + \frac{d\zeta_2}{dt}\Big|_{t=0}\cdot[F + \lambda G]_v\right\}dx = 0$$

for all admissible ζ_1 (ζ_2 is not arbitrary). Since (83) holds for all $\lambda(x)$, we will try to find a particular $\lambda(x)$ so that $[F + \lambda G]_v = 0$. Expanding, we have, since G does not contain v',

$$(84) \qquad [F + \lambda G]_v = F_v + \lambda G_v - \frac{d}{dx}F_{v'} = 0 \ .$$

Hence, we may solve for λ if $G_v \neq 0$. If $G_v = 0$, but $G_u \neq 0$, then the roles of ζ_1 and ζ_2 may be interchanged. If both vanish, the procedure breaks down. We note that since

$G_{v'} = 0$, the condition $G_v \neq 0$ may be written $[G]_v \neq 0$. This choice of λ leaves only the first term in (83),

$$\int_a^b \eta_1 [F + \lambda G]_u dx = 0 .$$

Therefore, we conclude that

$$[F + \lambda G]_u = 0 .$$

Hence, the Euler-Lagrange rule still holds, except that λ is now a function of x: i.e. we have shown that if u and v extremize $I(\Phi, \Psi)$ subject to $G(x,\Phi,\Psi) = 0$, then there is a function $\lambda(x)$, such that

$$[F + \lambda G]_u = [F + \lambda G]_v = 0 ,$$

provided that $[G]_u$ and $[G]_v$ do not both vanish. In general the Euler equations in u and v (if there are more dependent variables, there will be a corresponding number of Euler equations) are second order differential equations with fixed or natural end conditions: their solution will be in the form

$$u = u(x,\lambda(x))$$
$$v = v(x,\lambda(x)) .$$

The function $\lambda(x)$ is then determined by solving the equation $G(x,u,v) = 0$ for $\lambda(x)$.

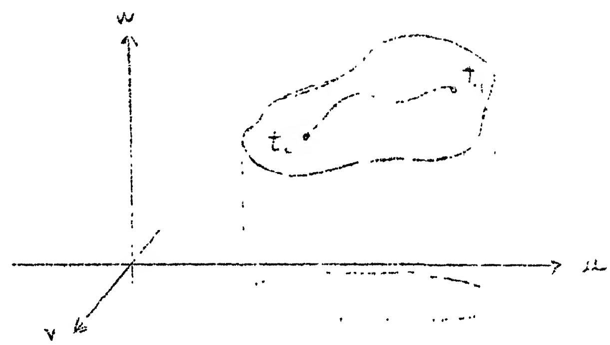
An example of this type of problem is that of geodesics on a surface. Suppose (u,v,w) are the rectangular coordinates of a point on a surface determined by the equation

(85) $$G(u,v,w) =$$

Then a curve on this surface may be given in the form

$$u = u(t), \quad v = v(t), \quad w = w(t),$$

where u, v, and w are not independent, but must satisfy (85). The arc length of the curve, which is defined to be a minimum between two fixed points for the geodesics is given by



(86) $$\int_{t_o}^{t_1} F dt = \int_{t_o}^{t_1} \sqrt{\dot{u}^2 + \dot{v}^2 + \dot{w}^2}\, dt \ .$$

Hence, we seek to minimize (86) subject to (85). The Euler equations are

$$\frac{d}{dt} \frac{\dot{u}}{\sqrt{\dot{u}^2 + \dot{v}^2 + \dot{w}^2}} - \lambda G_u = 0 \ ,$$

and similarly for v and w. Since the parameter t is arbitrary we may choose t = sa, where s is the arc length; then

$$\sqrt{\dot{u}^2 + \dot{v}^2 + \dot{w}^2} = \frac{1}{a} \ ,$$

and the equations become (with a different λ)

$$u'' - \lambda G_u = 0$$
$$v'' - \lambda G_v = 0$$
$$w'' - \lambda G_w = 0 \ .$$

As soon as a specific G is given, these together with G = 0 may be solved for u, v, w, and λ. We observe, however, that for any G, the geodesics must be such that the directions

$$G_u : G_v : G_w \ , \qquad u'' , v'', w''$$

are parallel. Since these are the directions respectively, of the normal to the surface and the principal normal to the curve, we conclude that the principal normal to a geodesic at every point coincides with the normal to the surface.

c) We finally consider the isoperimetric problem having a differential equation as a side condition. We seek to extremize

subject to the side condition

$$(87) \qquad G(x,u,u', v, v') = 0 \ .$$

This type of problem is encountered, for example, in non-holonomic dynamical systems, in which the number of independent coordinates is not equal to the number of degrees of freedom. The Euler-Lagrange rule still holds for this case, i.e. if u and v extremize $I(u, v)$ subject to (87) and if $[G]_u$, $[G]_v$ are not both zero, then there is a function $\lambda(x)$ such that

$$[F + \lambda G]_u = [F + \lambda G]_v = 0$$

and $G(x,u,v,u',v') = 0$.

The proof is more involved here because not even one of the dependent functions, say Φ, can be subjected to an arbitrary admissible variation. For, if we attempt this, Ψ is determined as the solution of a first order differential equation, but must satisfy two boundary conditions, which is in general impossible. This difficulty may be overcome be considering variations of the form $\Phi = u + t_1\zeta_1 + t_2\zeta_2$, $\Psi = v + \eta(t_1,t_2,x)$ where a relation between t_1 and t_2

is fixed in order to satisfy the second boundary condition.*

As an example, we consider the simplest problem of the first section only expressed in a slightly different manner. We seek to extremize

$$I = \int_a^b F(x,v,u)dx$$

subject to the condition

$$u = v' .$$

Here $F = F(x,v,u)$ and $G = u - v'$. Then, according to the rule, there should be a function $\lambda(x)$ so that $u - v' = 0$, and

$$[F + \lambda G]_u = [F + \lambda G]_v = 0 .$$

We have, respectively

$$F_u + \lambda = 0$$
$$F_v - \frac{d}{dx}\lambda(-1) = F_v + \lambda' = 0 .$$

Hence, $\lambda = - F_u$. Eliminating λ, we have

$$F_v - \frac{d}{dx} F_u = 0 .$$

or, since $u = v'$

$$F_v - \frac{d}{dx} F_{v'} = 0 ,$$

the usual Euler equation. Of course, this problem avoids the difficulty inherent in the general problems since only v (and not u) must satisfy boundary conditions, so that in this case arbitrary variations of v are permissible.

* For details, see Bolza, "Variationsrechnung", p. 558; or Hilbert, "Zur Variationrechnung", Math. Annal. Vol. LXII, No.3.

The problem of differential equation side conditions for the case of more than one independent variable is even more difficult, and has only been solved in special instances.

Problems

1) By use of the Euler-Lagrange rule, solve the isoperimetric problem of the circle--i.e. minimize

$$L(u) = \int_a^b \sqrt{1 + \dot{u}^2}\, dx$$

subject to the condition that

$$A(u) = \int_a^b u\,dx = A_0, \text{ a constant.}$$

2) Use the Euler-Lagrange rule to find the shape of a hanging chain (see page 37).

3) Find the geodesics on the sphere

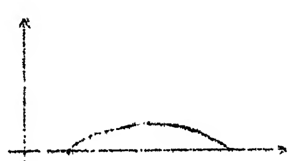
$$G(u,v,w) = u^2 + v^2 + w^2 - 1 = 0 .$$

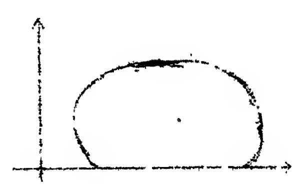
See page 41.

6. Parametric Form of the Theory. For many types of problems having physical origins the class of admissible functions considered up to now is too restricted. For example, in the iso-perimetric problem we seek a curve of given length which maximizes the area enclosed between the curve and the straight line joining its endpoints. If the given length is sufficiently small,

$$L < \pi(b - a)$$

the maximizing arc may be written in the form y where $a \le x \le b$. However if $L > \pi(b - a)/2$, the problem still





has a solution, but it is no longer expressible in the form of a function $y = u(x)$, $a \leq x \leq b$. This artificial restriction may be removed by expressing F in parametric form, as follows. Consider the functional

$$I = \int_{t_0}^{t_1} H(x,y,\dot{x},\dot{y})\,dt \;, \tag{88}$$

where $x = x(t)$, $y = y(t)$ are parametric equations of an admissible curve. We require that for an admissible curve

$$\begin{aligned} x(t_0) &= a, & x(t_1) &= b \\ y(t_0) &= \;\;, & y(t_1) &= B \end{aligned}$$

and that $x(t)$ and $y(t)$ are continuous functions of t, with piecewise continuous derivatives such that $\dot{x}^2 + \dot{y}^2 \neq 0$. The important point in this case is that for the problem to make sense H cannot be arbitrary but must be a homogeneous function of the first degree in $\dot{x}$ and $\dot{y}$. This follows from the requirement that I should depend only upon the curve joining the fixed end points and not upon the particular parametric representation used to describe that curve. Hence, if we replace the parameter t by another parameter $\tau = \tau(t)$ in a one to one way ($\dot{\tau} > 0$), I should not change. We therefore have the following equation:

$$\int_{t_0}^{t} H(x,y,\dot{x},\dot{y})\,dt = \int_{\tau_0}^{\tau} H\left(x,y,\dot{\tau}\frac{dx}{d\tau},\dot{\tau}\frac{dy}{d\tau}\right)\frac{d\tau}{\dot{\tau}}$$

$$= \int_{\tau_0}^{\tau} H\left(x,y,\frac{dx}{d\tau},\frac{dy}{d\tau}\right)d\tau \,.$$

This requires that, for $\dot{\tau} > 0$

$$H\left(x,y,\dot{\tau}\frac{dx}{d\tau},\dot{\tau}\frac{dy}{d\tau}\right) = \dot{\tau}\,H\left(x,y,\frac{dx}{d\tau},\frac{dy}{d\tau}\right) \,. \tag{89}$$

In particular if $\dot{\tau} = k$, a positive constant we have

$$H(x,y,k\dot{x},k\dot{y}) = kH(x,y,\dot{x},\dot{y}) ,$$

which is the definition of a homogeneous function of the first degree in $\dot{x}$ and $\dot{y}$. Thus, if a problem be expressed in parametric form, the integrand must be homogeneous of first degree in $\dot{x}$ and $\dot{y}$.

If $\dot{x} > 0$ throughout the interval (t_0, t_1), we may take $\dot{\tau} = 1/\dot{x}$ in (89), and setting $y' = \dot{y}/\dot{x}$ (89) becomes

$$H(x,y,\dot{x},\dot{y}) = \dot{x}H(x,y,1,y') = \frac{dx}{dt} F(x,y,y') ,$$

and we may write

$$(90) \qquad I = \int_{t_0}^{t_1} H(x,y,\dot{x},\dot{y})dt = \int_a^b F(x,y,y')dx .$$

Thus if x always increases as the curve is traversed (which is the case if $\dot{x} > 0$), the homogeneous problem may be transformed back into the inhomogeneous problem. Of course everything which has been done for $\dot{\tau} > 0$ and $\dot{x} > 0$ applies equally well if the direction of the inequality is reversed.

The integral (88) may be treated exactly as in the other problems we have considered. We obtain, as a necessary condition satisfied by an extremal $x = x(t)$ $y = y(t)$, the pair of simultaneous equations

$$(91) \qquad \begin{aligned} H_x - \frac{d}{dt} H_{\dot{x}} &= 0 \\ H_y - \frac{d}{dt} H_{\dot{y}} &= 0 . \end{aligned}$$

These equations must hold for the extremizing curve independently of the choice of parameter in terms of which it is defined. The choice of parameter is thus ... to expediency in any actual problem under ...sideration.

As an example, we consider the problem of determining the geodesics on a surface, phrased in parametric form rather than isoperimetrically (cf. page 40). Suppose the surface $G(u,v,w) = 0$ is defined parametrically as $u = u(x,y)$, $v = v(x,y)$, $w = w(x,y)$. That is, we have a correspondence between a region in the x,y plane and the surface G in the u,v,w space. A curve on G will then be represented by a curve in the x,y plane. If this curve be given parametrically in the form $x = x(t)$, $y = y(t)$, then the curve on the surface $G = 0$ will take the form $u = u(x(t),y(t))$, $v = v(x(t),y(t))$, $w = w(x(t),y(t))$. It has for element of arc length

$$ds = \sqrt{E\dot{x}^2 + 2F\dot{x}\dot{y} + G\dot{y}^2}\, dt \quad ,$$

where

$$E = u_x^2 + v_x^2 + w_x^2$$

$$F = u_x u_y + v_x v_y \quad w_x w_y$$

$$G = u_y^2 + v_y^2 + w_y^2$$

are given functions of x and y, and

$$\Delta = EG - F^2 > 0 \quad .$$

Denoting the radical by H, the integral to be minimized then becomes

$$I = \int_{t_o}^{t_1} H dt \quad .$$

Euler's equations (91) for the minimizing curve become

(92)

$$\frac{E_x\dot{x}^2 + 2F_x\dot{x}\dot{y} + G_x\dot{y}^2}{2H} - \frac{d}{dt}\frac{E\dot{x} + F\dot{y}}{H} = 0$$

$$\frac{E_y\dot{x}^2 + 2F_y\dot{x}\dot{y} + G_y\dot{y}^2}{2H} - \frac{d}{dt}\frac{F\dot{x} + G\dot{y}}{H} = 0 \quad .$$

In general (92) constitutes a very unwieldy system of equations. However, we recall that any parameter may be used in place of t. In particular, if we use a parameter s, proportional to the arc length, H is a constant and the equations (92) reduce to

$$
(93)\qquad
\begin{aligned}
E_x x'^2 + 2F_x x'y' + G_x y'^2 &= 2\,\frac{d}{ds}\,(Ex' + Fy')\\
E_y x'^2 + 2F_y x'y' + G_y y'^2 &= 2\,\frac{d}{ds}\,(Fx' + Gy')
\end{aligned}
$$

where $x' = \frac{dx}{ds}$, $y' = \frac{dy}{ds}$. These equations are then the appropriate system to use in determining the geodesics for any particular surface determined by the functions u, v, w.

We remark that although the parametric formulation of the problem results in two Euler equations (91), these equations are not independent because of the requirement that H be homogeneous of first degree in $\dot x$ and $\dot y$. (This, of course, is to be expected, since an equivalent inhomogeneous problem would have only one Euler equation.) In fact, H must satisfy the Euler homogeneity identity

$$
(94)\qquad H = \dot x H_{\dot x} + \dot y H_{\dot y}
$$

and combining (91) with (94) we obtain the single equation

$$
(95)\qquad H_{x\dot y} - H_{y\dot x} + \bar H(\dot x\ddot y - \ddot x\dot y) = 0 ,
$$

where

$$
\bar H = \frac{H_{\dot x\dot x}}{\dot y^2} = \frac{H_{\dot x\dot y}}{-\dot x\dot y} = \frac{H_{\dot y\dot y}}{\dot x^2} ,
$$

which is equivalent to the system (91).

The homogeneous formulation may be carried out for the case of more than one independent variable in a straightforward way, the only formal change being that the Euler equations will be partial differential equations. We

consider as an example the Plateau problem of minimal surfaces already discussed on page 25. We will now formulate this problem parametrically.

Let $x = x(t_1,t_2)$, $y = y(t_1,t_2)$, $z = z(t_1,t_2)$ be the parametric equations of a surface such that as t_1,t_2 traces out the curve g in the t_1,t_2 plane x,y,z traces out the fixed space curve Γ in the x,y,z space thus satisfying the boundary condition. We wish to find the functions x,y,z for which the area of the surface enclosed by Γ is a minimum--i.e. to minimize the integral

$$I = \iint_g \sqrt{EG - F^2}\, dt_1 dt_2 = \iint_g \sqrt{W}\, dt_1 dt_2$$

where

$$E = x_{t_1}^2 + y_{t_1}^2 + z_{t_1}^2$$

$$F = x_{t_1}\cdot x_{t_2} + y_{t_1}\cdot y_{t_2} + z_{t_1}\cdot z_{t_2}$$

$$G = x_{t_2}^2 + y_{t_2}^2 + z_{t_2}^2 \, .$$

The resulting Euler equations are

$$\frac{\partial}{\partial t_1}\left(\frac{W_{x_{t_1}}}{\sqrt{W}}\right) + \frac{\partial}{\partial t_2}\left(\frac{W_{x_{t_2}}}{\sqrt{W}}\right) = 0$$

$$\frac{\partial}{\partial t_1}\left(\frac{W_{x_{t_1}}}{\sqrt{W}}\right) + \frac{\partial}{\partial t_2}\left(\frac{W_{x_{t_2}}}{\sqrt{W}}\right) = 0$$

$$\frac{\partial}{\partial t_1}\left(\frac{W_{x_{t_1}}}{\sqrt{W}}\right) + \frac{\partial}{\partial t_2}\left(\frac{W_{x_{t_2}}}{\sqrt{W}}\right) = 0 \, .$$

Here as before the equations can be greatly simplified by a proper choice of parameters. We can always choose t_1 and t_2 so that $F = 0$ and $E = G$, and with this choice, equations (95) reduce to the very elegant form

$$
\begin{aligned}
x_{uu} + x_{vv} &= 0 \\
y_{uu} + y_{vv} &= 0 \qquad (97) \\
z_{uu} + z_{vv} &= 0
\end{aligned}
$$

Problems

1) Verify that (97) results from (96) on setting $F = 0$, $E = G$. Compare with the previous method used, page 23.
2) Use (93) to find the geodesics on a sphere, representing the sphere parametrically by

$$
\begin{aligned}
u &= \sin x \sin y \\
v &= \sin x \cos y \\
w &= \cos x .
\end{aligned}
$$

Compare with the previous methods mentioned, i.e. in simplest form (page 11) and in isoperimetric form (page 40).
3) Derive (95) from (91) and (94).

7. Invariance of the Euler Equation. On page 13 we mentioned that the Euler Equation may be thought of as a generalisation of the vector equation grad $f = 0$, where $f = f(x_1,\ldots x_n)$. We recall that if the independent variables $x_1,\ldots x_n$ are transformed into new variables $\xi_1(x_1,\ldots x_n),\ldots\xi_n(x_1,\ldots x_n)$ in such a way that the Jacobian

$$\frac{\partial(x_1,\ldots x_n)}{\partial(\xi_1,\ldots\xi_n)} \neq 0 \text{ or } \infty,$$

then if $\text{grad}_{(x)} f = 0$ at a point $(\bar{x}_1,\ldots\bar{x}_n)$ it will follow that $\text{grad}_{(\xi)} f = 0$ at the corresponding point $(\bar{\xi}_1,\ldots\bar{\xi}_n)$. We say that the equation grad $f = 0$ is invariant under transformation of the coordinate system. We would then expect that the Euler equation should also be invariant

under transformation of the coordinate system. That this is so may be verified by considering the result of replacing x by $\xi = \xi(x)$ in

$$I = \int_{x_0}^{x_1} F(x,u,u_x)dx$$

$$= \int_{\xi_0}^{\xi_1} F[x(\xi),u(x(\xi)),\ u_\xi(x(\xi))\xi_x]x_\xi d\xi$$

$$= \int_{\xi_0}^{\xi_1} H(\xi,v,v_\xi)d\xi \ .$$

where $v(\xi) = u(x(\xi))$. A one parameter family $u(x) + t\zeta(x)$ will correspond to $v(\xi) + t\eta(\xi)$ where $\zeta(x(\xi)) = \eta(\xi)$; substituting these expressions into the above integrals, differentiating with respect to t and setting $t = 0$, we have

$$\int_{x_0}^{x_1} \zeta[F]_{u(x)}dx = \int_{\xi_0}^{\xi_1} \eta\ [H]_{v(\xi)}d\xi = 0 \ .$$

Since $d\xi = \xi_x dx$ and $\zeta(x) = \eta(\xi)$, we conclude that

$$\int_{x_0}^{x_1} \zeta \left\{ [F]_u - [H]_v\xi_x \right\} dx = 0$$

or

$$[F]_u = [H]_v\xi_x \ . \tag{98}$$

In the case of more than one independent variable, a change of coordinate system from $(x_1,\ldots x_n)$ to $(\xi_1,\ldots,\xi_n)$ requires that the Euler operator in the new system be multiplied by the Jacobian of the transformation. We have

$$[F(x_1,\ldots x_n;\ u,u_{x_1},\ldots u_{x_n})]_u$$

$$= [H(\xi_1,\ldots\xi_n;\ v,v_{\xi_1},\ldots v_{\xi_n})]_v\frac{\partial(\xi_1,\ldots\xi_n)}{\partial(x_1,\ldots,x_n)}$$

where

$$v(\xi_1,\dots\xi_n) = u(x_1(\xi_1,\dots\xi_n),\dots x_n(\xi_1,\dots\xi_n))$$

and

$$H = F[x_1(\xi_1,\dots\xi_n),\dots x_n(\xi_1\dots\xi_n);\ v(\xi_1,\dots\xi_n);\ u_{x_1}(\xi_1,\dots\xi_n),$$
$$\dots u_{x_n}(\xi_1\dots\xi_n)]\frac{\partial(x_1,\dots x_n)}{\partial(\xi_1,\dots\xi_n)}$$

with

$$u_{x_i} = v_{\xi_1}\frac{\partial\xi_1}{\partial x_i} + \dots + v_{\xi_n}\frac{\partial\xi_n}{\partial x_i} .$$

In particular, if for a given u, $[F]_u = 0$, then $[H]_v = 0$, so that the Euler equation is invariant. The property that a curve be an extremal (i.e. a solution of Euler's equation) remains unaltered by a transformation of the independent variables.

The invariance of the Euler expression is a useful principle in actually computing the transformed differential expression. One of the most important differential operators is the Laplacian $\Delta u = u_{xx} + u_{yy} = -\frac{1}{2}[F]_u$, where $F = u_x^2 + u_y^2$. If we wish to find Δu in polar coordinates $x = r\cos\theta$, $y = r\sin\theta$, we may calculate $\frac{\partial(x,y)}{\partial(r,\theta)} = r$, $H = r(v_r^2 + \frac{1}{r^2}v_\theta^2)$, and conclude that

$$[F]_u = [H]_v\cdot\frac{1}{r} = \frac{1}{r}(H_v - \frac{\partial}{\partial r}H_{v_r} - \frac{\partial}{\partial\theta}H_{v_\theta})$$

$$= -\frac{2}{r}(rv_{rr} + v_r + \frac{1}{r}v_{\theta\theta}) .$$

Hence, from $u = -\frac{1}{2}[F]_u$

$$\Delta v = v_{rr} + \frac{1}{r} v_r + \frac{1}{r^2} v_{\theta\theta} ,$$

a result obtained without computation of second derivatives.

Problems

1) Find Δu for spherical coordinates $x = r \cos\theta \sin\psi$, $y = r \sin\theta \sin\psi$, $z = r \cos\psi$. Generalize to n dimensions.
2) Find Δu for $\xi = \xi(x,y)$, $\eta = \eta(x,y)$.
3) In n-dimensions,

$$\Delta u = \sum_{i=1}^{n} u_{x_i x_i} .$$

Let $x_i = x_i(\xi_1,\ldots\xi_n)$ and $\xi_i = \xi_i(x_1,\ldots x_n)$, and define

$$g^{ik} = \sum_{j=1}^{n} \frac{\partial \xi_i}{\partial x_j} \cdot \frac{\partial \xi_k}{\partial x_j} .$$

Then show that

(a) $$D = \begin{vmatrix} g^{11} & g^{12} \ldots g^{1n} \\ g^{21} & \vdots \\ g^{n1} & \ldots\ g^{nn} \end{vmatrix} = \left[\frac{\partial(\xi_1,\ldots\xi_n)}{\partial(x_1,\ldots x_n)}\right]^2$$

(b) $$\Delta u = \sqrt{D} \cdot \sum_j \frac{\partial}{\partial \xi_j} \left(\frac{1}{\sqrt{D}} \sum_k g^{jk} u_{\xi_k}\right) .$$

8. The Legendre Condition. In the theory of extrema of functions of a single variable, a necessary condition for a minimum, besides $f'(x) = 0$, is that $f'' \geq 0$ (if it exists) A condition somewhate analogous to this will be seen to hold in Calculus of Variations.

Let us suppose that there is an admissible function u, for which

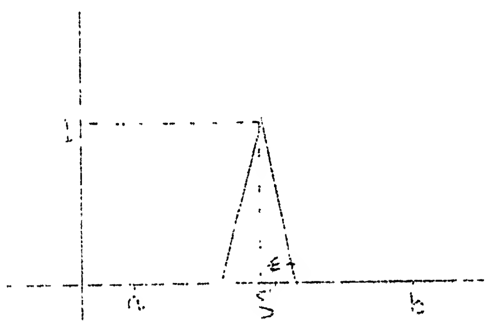
$$I(u) = \int_a^b F(x,u,u')dx$$

is a minimum. Then $G(t) = I(u + t\zeta)$ has a minimum at $t = 0$: accordingly $G'(0) = 0$ (from which follows the Euler equation $[F]_u = 0$) and also $G''(0) \geq 0$, assuming its existence. Hence for every admissible ζ,

$$(99) \qquad G''(0) = \int_a^b [F_{uu}\zeta^2 + F_{uu'}\zeta\zeta' + F_{u'u'}\zeta^2]dx \geq 0 .$$

We choose a special variation

$$\zeta(x) = \begin{cases} 0 & a \leq x \leq \xi - \varepsilon \\ 1 + (x - \xi)/\varepsilon & \xi - \varepsilon \leq x \leq \xi \\ 1 - (x - \xi)/\varepsilon & \xi \leq x \leq \xi + \varepsilon \\ 0 & \xi + \varepsilon \leq x \leq b . \end{cases}$$



If we substitute this function in (99) and let $\varepsilon \to 0$, the term

$$\frac{1}{\varepsilon^2} \int_{\xi-\varepsilon}^{\xi+\varepsilon} F_{u'u'}dx$$

will dominate the left side of the inequality (99) and determine its sign. Thus the sign of $F_{u'u'}$ determines the sign of $G''(0)$, and for a minimum, the weak Legendre Condition

$$(100) \qquad F_{u'u'} \geq 0 .$$

must hold for all points on the curve u. We have seen (page 33) that $F_{u'u'} \neq 0$ is essential in order that the Euler equation should not degenerate. In many problems the strong Legendre condition $F_{u'u'} > 0$ holds, but this is still not <u>sufficient</u> to guarantee a minimum. (We will see in the next chapter that if $F_{\Phi'\Phi'} > 0$ for all admissible Φ, then a solution of the Euler equation is a minimum.)

Problems frequently take the form $F = g(\Phi)\sqrt{1 + \Phi'^2}$ for which $F_{\Phi'\Phi'} = g/(1 + \Phi'^2)$ and $F_{\Phi'\Phi'} > 0$ if $g(\Phi) > 0$, so that for such a function Legendre's condition is sufficient.

In case F contains more dependent functions, the Legendre condition is that the matrix

$$(F_{u_i'u_j'})$$

be positive definite, that is

$$\sum_i \sum_j \lambda_i\lambda_j F_{u_i'u_j'} \geq 0$$

for all λ_i, λ_j.

Problems

1) Verify the last statement above.

II HAMILTON JACOBI THEORY - SUFFICIENT CONDITIONS

1. The Legendre Transformation. Transformation from point to line coordinates is frequently useful in the theory of differential equations. We consider first the one dimensional case of a curve $u = f(x)$, which can be considered as the envelope of its tangent lines. The tangent line at any point x,u is given by

(1) $$U + xf'(x) - f(x) = f'(x)X ,$$

U and X being coordinates along the line. The line (1) is determined by its coordinates (i.e. slope and intercept)

(2) $$\begin{cases} \xi = f'(x) \\ w = xf'(x) - u \end{cases}$$

where a unique value of (ξ,w) is assigned to each point (x,u). The curve $u = f(x)$ can then be represented as $w = W(\xi)$ on elimination of x between the equations (2), which can always be done if $f''(x) \neq 0$.* Between corresponding sets of coordinates (x,u) and (ξ,w) (i.e. referring to the same point on the curve) there exists the symmetric relation

(3) $$u + w = x\xi$$

in virtue of (1).

The equations (2) allow a transformation from point to line coordinates. To reverse this procedure we suppose $w = W(\xi)$ is given and find the envelope of the one parameter family of lines

---

*If $f''(x) \neq 0$, $\xi = f'(x)$ can be solved for x, and substitution in $w = xf'(x) - f(x)$ gives the required relation. Inversion of $w = xf'(x) - f(x)$ instead of $\xi = f'(x)$ requires $xf''(x) \neq 0$ so nothing is gained. The points $f''(x) = 0$ (inflection points) are singularities in line coordinates, and $f''(x) \equiv 0$ represents a single straight line. Similarly $W''(\xi) = 0$ is a cusp in point coordinates, and $W''(\xi) \equiv 0$ represents a pencil of lines through a fixed point (no envelope). Duality is observed as points on a line ($f''(x) \equiv 0$) and lines through a point ($W''(\xi) \equiv 0$).

II HAMILTON JACOBI THEORY - SUFFICIENT CONDITIONS

1. Legendre Transformation. Transformation from point to line coordinates is frequently useful in the theory of differential equations. We consider first the one dimensional case of a curve $u = f(x)$, which can be considered as the envelope of its tangent lines. The tangent line at any point x,u is given by

(1) $$U + xf'(x) - f(x) = f'(x)X ,$$

U and λ being coordinates along the line. The line (1) is determined by its coordinates (i.e. slope and intercept)

(2) $$\begin{cases} \xi = f'(x) \\ w = xf'(x) - u \end{cases}$$

where a unique value of (ξ,w) is assigned to each point (x,u). The curve $u = f(x)$ can then be represented as $w = W(\xi)$ on elimination of x between the equations (2), which can always be done if $f''(x) \neq 0$.* Between corresponding sets of coordinates (x,u) and (ξ,w) (i.e. referring to the same point on the curve) there exists the symmetric relation

(3) $$u + w = x\xi$$

in virtue of (1).

The equations (2) allow a transformation from point to line coordinates. To reverse this procedure we suppose $w = W(\xi)$ is given and find the envelope of the one parameter family of lines

---

*If $f''(x) \neq 0$, $\xi = f'(x)$ can be solved for x, and substitution in $w = xf'(x) - f(x)$ gives the required relation. Inversion of $w = xf'(x) - f(x)$ instead of $\xi = f'(x)$ requires $xf''(x) \neq 0$ so nothing is gained. The points $f''(x) = 0$ (inflection points) are singularities in line coordinates, and $f''(x) \equiv 0$ represents a single straight line. Similarly $W''(\xi) = 0$ is a in point coordinates, and $W''(\xi) \equiv 0$ represents a pencil of lines through a fixed point (no envelope). Duality is observed as points on a line ($f''(x) \equiv 0$) and lines through a point ($W''(\xi) \equiv 0$).

$$U + w(\xi) = \xi\lambda \; . \tag{4}$$

To do this we differentiate with respect to the parameter ξ and combine with (4), setting $U = u$ and $\lambda = x$ (coordinates along the envelope), obtaining

$$\begin{cases} x = w'(\xi) \\ u = \xi w'(\xi) - w \; . \end{cases} \tag{5}$$

The duality of the transformation is evident on comparison of (2) and (5). Elimination of ξ in (5) to obtain $u = f(x)$ is possible if $w''(\xi) \neq 0$.

Another more formal way of deriving (5) by inversion of (2) is to differentiate equation (3) with respect to ξ treating u,w and x as functions of ξ. We have

$$f'(x)\,\frac{dx}{d\xi} + w'(\xi) = \xi\,\frac{dx}{d\xi} + x \; .$$

But, since $\xi = f'(x)$ from (2), this reduces to

$$x = w'(\xi) \; ,$$

and the other part of (5) is obtained by substitution in (3).

The foregoing can be easily extended to the case of n independent variables. We have $u = f(x_1, \ldots x_n)$ and introduce as coordinates the direction numbers and intercept of the tangent plane to this surface

$$\begin{cases} \xi_i = f_{x_i} \\ w = \sum_{j=1}^{n} x_j f_{x_j} - f \; . \end{cases} \tag{6}$$

The parameters x_i can be eliminated if the Hessian $|f_{x_i x_j}| \neq 0$,* (i.e. the Jacobian of $\xi_i = f_{x_i}$ does not vanish) yielding

---

*The condition $|f_{x_i x_j}| \equiv 0$ means that the family of tangent planes is a smaller than n parameter family. For example, with n = 2 there is a one parameter family, i.e. a developable, or a zero parameter family, a plane.

$$w = W(\xi_1, \ldots \xi_n) .$$

As before, the symmetric relation

$$(7) \qquad u + w = \sum_{i=1}^{n} x_i \xi_i$$

holds, and differentiation with respect to ξ_j gives

$$\sum_{i=1}^{n} f_{x_i} \frac{\partial x_i}{\partial \xi_j} + W_{\xi_j} = x_j + \sum_{i=1}^{n} \frac{\partial x_i}{\partial \xi_j} \xi_i$$

which, on using $\xi_i = f_{x_i}$, and then substituting in (7) results in

$$(8) \qquad \begin{cases} x_j = W_{\xi_j} \\ u = \sum_{i=1}^{n} \xi_i W_{\xi_i} - W . \end{cases}$$

The Legendre transformation is most frequently useful in transforming partial differential equations in which derivatives appear in a more complicated form than the independent variables.

2. <u>The Distance Function - Reduction to Canonical Form.</u> In this chapter we will be concerned with the variational problem represented by the integral

$$(9) \qquad I(u_1, \ldots u_n) = \int_{\tau}^{t} F(s;\, u_1, \ldots u_n;\, u_1', \ldots u_n')\,ds$$

of which the Euler equations are

$$(10) \qquad F_{u_i} - \frac{d}{ds} F_{u_i'} = 0 .$$

The solutions of (10) (extremals) represent families of curves $u_i(s)$ in the n+1 dimensional space $(s, u_1, \ldots u_n)$, it being possible to impose on each curve 2n conditions (if we ignore degeneracy). Thus we can consider a single curve passing through each point (τ, k) of the space (i.e.

$s = \tau$, $u_i = k_i$) and satisfying there the initial conditions $u'_i = k'_i$, or consider an extremal to connect every pair of points (τ, k) (t, q) in some neighborhood (boundary value problem), or we can consider an n parameter family of extremals passing through the point (τ, k), leaving open the remaining n conditions. If we restrict the path of integration of (8) to be always taken along extremals then a unique "distance"* I is associated with every pair of points (τ, k), (t, q) which can be connected by an extremal, thus defining I as a function of the $2n + 2$ variables $(\tau;\ k_1, \dots k_n;\ t;\ q_1, \dots q_n)$:

(11) $$I(\tau, k, t, q) = \int_{\tau}^{t} F(s, u, u')ds \ .$$

An essential concept in the theory of sufficient conditions is that of a field which is defined as follows: a family of curves in a neighborhood is said to form a field if one and only one curve of the family passes through each point of the neighborhood. In particular an n parameter family of extremals through a point can form a field, and in general to form a field in n+1 dimensional space an n parameter family of curves is required. If we have a neighborhood covered by a field of extremals through a point, then by use of (11) it is possible to define a single valued function I over the neighborhood. In this case it will be convenient to work in the n+2 dimensional (I, s, u) space.

As an example consider

(12) $$I = \int_{\tau}^{t} \sqrt{1 + (\frac{du}{ds})^2}\, ds \ ,$$

for the length of a curve u(s), for which the Euler equation is $\frac{d^2u}{ds^2} = 0$. Here n = 1 and we have a two parameter family

---

*If F is a quadratic form in u'_i then the extremals are geodesics and I is actually distance. More generally F can be considered a metric in a "Finsler Space" with Finsler distance I.

of straight lines as extremals, joining every pair of points in the (s,u) plane. If (τ,k) is fixed we have a field consisting of all rays through (τ,k), and $I(t,q)$ is a cone with vertex at $(0,\tau,k)$ in the three dimensional (I,t,q) space.

If in (11) we hold (τ,k) fast, I is defined as a function of (t,q) in a neighborhood of (τ,k) and this function can then be considered to depend on the n+1 parameters (τ,k). By the usual method of elimination of parameters a partial differential equation satisfied by this n+1 parameter family of functions can be found, namely:

$$(13)\qquad \begin{cases} I_t = F(t,q,q') - \sum_{i=1}^{n} q_i' F_{q_i'}(t,q,q') \\ I_{q_i} = F_{q_i'}(t,q,q') \end{cases}$$

which is relation between the n+1 derivatives I_t, I_q expressed implicitly in terms of n parameters $q_1', \ldots q_n'$. We observe that the formalism for eliminating the parameters q' is exactly that of the Legendre Transformation (cf. (8)). Rather than proceeding in this way, we shall impose the Legendre transformation at the outset, in (9), thereby obtaining the reduction of (9) and (10) to canonical form. We observe, however, that the proper Legendre transformation is applied in terms of u' (i.e. q') which would necessitate subjecting u to a very complicated transformation. In order to avoid this difficulty we introduce the functions $p_i = u_i'$ and consider the variational problem

$$(14)\qquad I(u_1,\ldots u_n;\ p_1,\ldots p_n) = \int_{\tau}^{t} F(s:\ u_1,\ldots u_n;\ p_1,\ldots p_n)ds$$

subject to the side conditions

$$p_i = u_i' \ .$$

This suggests the use of the Lagrange multiplier rule by which (14) becomes

$$I(u,p) = \int_{\tau}^{t} [F(s,u,p) + \sum_{i=1}^{n} \lambda_i(u_i' - p_i)]ds$$

for which the Euler equations are

$$F_{u_i} - \frac{d}{ds}\lambda_i = 0, \quad F_{p_i} - \lambda_i = 0,$$

and elimination of λ_i finally leads us to consider

$$(15) \qquad I(u,p) = \int_{\tau}^{t} [F(s,u,p) + \sum_{i=1}^{n} F_{p_i}(u_i' - p_i)]ds ,$$

together with the necessary restriction on F that $|F_{p_ip_j}| \neq 0$. Since the Lagrange multiplier rule was not proved for differential equations as side conditions, we accept (15) and show independently that it is equivalent to (9). By equivalence we mean that, although the function space over which I(u,p) is defined is much wider (2n independent functions u and p rather than n functions u), the extremals for both integrals are the same. Writing down the Euler equations of (15)

$$F_{u_j} + \sum_{i=1}^{n} F_{p_iu_j}(u_i' - p_i) - \frac{d}{ds} F_{p_j} = 0$$

$$F_{p_j} + \sum_{i=1}^{n} F_{p_ip_j}(u_i' - p_i) - F_{p_j} = \sum_{i=1}^{n} F_{p_ip_j}(u_i' - p_i) = 0$$

But, since $|F_{p_ip_j}| \neq 0$, the second equation implies $u_i' - p_i = 0$, and substitution into the first yields

$$F_{u_j} - \frac{d}{ds} F_{u_j'} = 0 ,$$

which is the set of Euler equations for (9). It is interesting to note that th variational problem (15) is degenerate, 2n first order equations replacing n of the second order.

In (15) the u's and p's are independent, so the Legendre transformation

$$(16)\quad \begin{cases} v_i = F_{p_i}(s,u,p) \\ L(s,u,v) = \sum_{i=1}^{n} p_i v_i - F(s,u,p) \end{cases}$$

can be applied treating u and s as parameters. The dependence of L on s, u, and v is given by eliminating p from (16). The condition for elimination is satisfied in virtue of our assumption $|F_{p_i p_j}| \neq 0$. We now have

$$(17)\qquad I(u,v) = \int_{\tau}^{t} [\sum_{i=1}^{n} u_i' v_i - L(s,u,v)]ds$$

which is again a degenerate variational problem in the 2n functions $(u_1,\ldots u_n;\ v_1,\ldots v_n)$. The Euler equations of (17) take the canonical form

$$(18)\quad \begin{cases} \frac{d}{ds} v_i + L_{u_i} = 0 \\ \frac{d}{ds} u_i - L_{v_i} = 0\ . \end{cases}$$

These transformations have their origin in classical mechanics, the equations (10) being Lagrange's equations of motion for a conservative system in the generalized coordinates u_i and momenta p_i (or v_i), F being the difference between kinetic and potential energies, and L their sum. The canonical form (18) can of course be derived directly from (10) without reference to a variational integral.

3. <u>The Hamilton-Jacobi Partial Differential Equation.</u> We will now consider (17) in a neighborhood in which it is assumed that a unique extremal exists joining any two points (τ,k) (t,q), the conjugate function v taking on the values l and λ (i.e. $v_i = l_i$ at $s = t$, $v_i = \lambda_i$ at $s = \tau$). The extremal passing through these points can be represented by

$$(19)\quad \begin{cases} u_i = f_i(s,t,q,\tau,k) \\ v_i = g_i(s,t,q,\tau,k) \end{cases}$$

where

$$(20)\qquad \begin{cases} f_i(t,t,q,\tau,k) = q_i, & f_i(\tau,t,q,\tau,k) = k_i \\ g_i(t,t,q,\tau,k) = l_i, & g_i(\tau,t,q,\tau,k) = \lambda_i . \end{cases}$$

In order to find the partial differential equation satisfied by the function $I(t,q,\tau,k)$ we must calculate its partial derivatives, I_t, I_{q_i}, etc. and this can be done most concisely by use of the variational symbol δ. Let the independent variables be taken as functions of a parameter ε; viz. $t(\varepsilon)$, $\tau(\varepsilon)$, $q_i(\varepsilon)$, $k_i(\varepsilon)$, $l_i(\varepsilon)$, $\lambda_i(\varepsilon)$; I then becomes a function of ε. We have

$$\frac{dI}{d\varepsilon} = I_t \frac{dt}{d\varepsilon} + I_\tau \frac{d\tau}{d\varepsilon} + \sum_{i=1}^{n} I_{q_i} \frac{dq_i}{d\varepsilon} + \sum_{i=1}^{n} I_{k_i} \frac{dk_i}{d\varepsilon} ,$$

and letting $\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = \delta I$, $\left.\frac{dt}{d\varepsilon}\right|_{\varepsilon=0} = \delta t$, etc. we have

$$(21)\qquad \delta I = I_t\delta t + I_\tau \delta\tau + \sum_{i=1}^{n} I_{q_i}\delta q_i + \sum_{i=1}^{n} I_{k_i}\delta k_i .$$

Performing the variation in

$$I(\varepsilon) = \int_{\tau(\varepsilon)}^{t(\varepsilon)} [\sum u_i'(\varepsilon)\dot{v}_i(\varepsilon) - L(s,u(\varepsilon),v(\varepsilon))]ds$$

where

$$u_i(\varepsilon) = f_i(s,t(\varepsilon),q(\varepsilon),\tau(\varepsilon),k(\varepsilon))$$
$$v_i(\varepsilon) = g_i(s,t(\varepsilon),q(\varepsilon),\tau(\varepsilon),k(\varepsilon))$$

we get

$$\delta I = [\sum q_i' l_i - L(t,q,l)]\delta t - [\sum k_i'\lambda_i - L(\tau,k,\lambda)]\delta\tau$$
$$+ \int_\tau^t \sum_{i=1}^{n} (u_i'\delta v_i + v_i\delta u_i' - L_{u_i}\delta u_i - L_{v_i}\delta v_i)ds .$$

The terms $\sum(u_i' - L_{v_i})\delta v_i$ in the integral vanish by (18)

(u and v are extremals), and on integrating by parts the remaining terms in the integral become

$$[\sum v_i \delta u_i]_\tau^t - \int_\tau^t \sum (v_i' + L_{u_i})\delta u_i ds = \sum_{i=1}^n [v_i \delta u_i]_\tau^t .$$

We now evaluate δu at $s = t$ and $s = \tau$. This is not equal to δq or δk because in δq, the variation is performed after s is set equal to t, that is after s has become a function of ε. Differentiating (19) we have

(22) $$\delta u = f_t \delta t + f_q \delta q + f_\tau \delta \tau + f_k \delta k .^*$$

However, differentiating the first relation in (20),

(23) $$\delta q = f' \delta t + f_t \delta t + f_q \delta q + f_\tau \delta \tau + f_k \delta k ,$$

where f' is the derivative of f with respect to its first argument, evaluated at $s = t$: i.e. $f' = q'$. By equating coefficients of the various δ's in (23) we see that $f_t = -q'$, $f_q = 1$, $f_\tau = f_k = 0$ evaluated at $s = t$. Similarly $f_\tau = -k'$, $f_k = 1$, $f_t = f_q = 0$ at $s = \tau$, and (22) reduces to

$$\delta u|_{s=t} = \delta q - q' \delta t$$
$$\delta u|_{s=\tau} = \delta k - k' \delta \tau ,$$

from which we get

$$\sum_{i=1}^n [v_i \delta u_i]_\tau^t = \sum (l_i \delta q_i - l_i q_i' \delta t - \lambda_i \delta k_i + \lambda_i k_i' \delta \tau)$$

and finally

(24) $$\delta I = \sum_{i=1}^n l_i \delta q_i - L(t,q,l)\delta t - \sum_{i=1}^n \lambda_i \delta k_i + L(\tau,k,\lambda)\delta\tau .$$

We now read off

---

*The subscript i is omitted.

$$(25)\quad \begin{cases} I_t = -L(t,q,l)\ , & I_{q_i} = l_i \\ I_\tau = L(\tau,k,\lambda)\ , & I_{k_i} = -\lambda_i\ . \end{cases}$$

In particular if we assume the endpoint (τ,k) is fixed, I as a function of t and q satisfies the first order Hamilton-Jacobi partial differential equation

$$(26)\qquad I_t + L(t;\ q_1,\ldots q_n:\ I_{q_1},\ldots I_{q_n}) = 0$$

obtained by elimination of the l_i from (25).

<u>Problem</u>

Evaluate the partial derivatives (25) by direct differentiation rather than variation.

The equation (26) was derived from a function I having n+1 parameters, but its importance lies in the existence of other more general solutions, depending on arbitrary functions rather than on a number of parameters. These solutions will be shown to represent distances measured from an initial surface rather than from an initial point as above. In order to motivate the manipulation to follow, we will first discuss briefly the general theory of partial differential equations of the first order, proofs being given only when necessary for our application.

First of all, it is clear that if from an n parameter family of solutions of any partial differential equation we form an n-1 parameter family by considering one of the parameters a_n as a function of the others, $a_n = f(a_1,\ldots a_{n-1})$, the envelope of this solution with respect to its n-1 parameters is also a solution (since it has at every point the same tangent plane as a known solution), and in fact depends on an arbitrary function, namely f.

Now consider the simple case of a first order quasi-linear partial differential equation in two independent variables

$$au_x + bu_y = c \ , \tag{27}$$

a, b and c being functions of x, y and u. Geometrically (27) states that at every point of a solution $u(x,y)$ the normal to the surface $u = u(x,y)$ (direction components $p = u_x$, $q = u_y$, -1) is perpendicular to the line element having components (a,b,c) at this particular point. In other words the element of surface contains the line element (a,b,c). If we integrate the ordinary differential equations

$$\frac{dx}{ds} = a(x,y,u) \ , \quad \frac{dy}{ds} = b(x,y,u) \ , \quad \frac{dz}{ds} = c(x,y,u) \tag{28}$$

we get a two parameter field of <u>characteristic curves</u> in the x,y,u space. Any solution of (27) is a one parameter family of characteristics, and conversely. Given a non-characteristic initial curve $x = x(s)$, $y = y(s)$, $u = u(s)$, the family of characteristics intersecting this line traces out an integral surface. In this way we have reduced the problem of the solution of the partial differential equation (27) to that of the integration of the three ordinary differential equations (28). It is clear that two integral surfaces can intersect only along characteristics, since to any common point belongs the whole characteristic through that point, and conversely this property completely characterizes the characteristic curves.

Let us now consider the more general equation

$$F(x,y,u,p,q) = 0 \ . \tag{29}$$

We note that for a linear equation (a,b,c in (27) functions of x and y alone), the characteristics are independent of u and form a one parameter family in the x,y plane. From this we would expect that the characteristics in the general case (29) would depend not only on u but also on p and q, and would not be fixed curves in the x,y,u space. In order to reduce this problem to that of ordinary differential equations it would therefore seem reasonable to consider curves $x(s)$, $y(s)$, $u(s)$, $p(s)$, $q(s)$ in five dimensions. In the x,y,u space

such a curve represents a space curve $x(s)$, $y(s)$, $u(s)$ together with a surface element $p(s)$, $q(s)$ associated with each point. At a given point (x,y,u) (29) allows a one parameter relation between p and q. Geometrically this means the surface element (tangent plane) traces out a one parameter family passing through the point, and envelopes a cone (in the quasilinear case the cone degenerates into a line, with a one parameter family of planes through the line). The line elements of these cones are the characteristic directions and are given by the differential equations*

$$(30) \qquad \frac{dx}{ds} = F_p \ , \qquad \frac{dy}{ds} = F_q , \qquad \frac{du}{ds} = pF_p + qF_q \ .$$

A curve which is a solution of the system (29) (subject to (30)) has a characteristic direction at every point, but a family of such curves is not necessarily an integral of (29). In order that these curves lie on an integral surface p and q must satisfy the two further relations

$$(31) \qquad \frac{dp}{ds} = - (pF_u + F_x) \qquad \frac{dq}{ds} = - (qF_u + F_y) \ .$$

Using (30) and (31) we have five ordinary differential equations in the five variables $x(s)$, $y(s)$, $u(s)$, $p(s)$, $q(s)$. In general a solution curve will exist passing through an arbitrary initial point x_o, y_o, u_o, p_o, q_o. There is a five parameter family of solutions, but since s admits of an arbitrary translation, and (29) is a relation between the initial values, the solutions reduce to three parameters. Interpreted in the (x,y,u) space, an initial value is a point (x,y,u) and a surface element (p,q), where (p,q) is not arbitrary but must satisfy (29) (i.e. must be tangent to a cone). Given such an initial value a characteristic curve $x(s)$, $y(s)$, $u(s)$ is determined, and with it tangent elements $p(s)$, $q(s)$ comprising characteristic strip. As in the quasilinear case integral ces are composed of one parameter ies of characteristic curves or strips. Given an initial curve $x(s)$, $y(s)$, s) we use the relation

*For proofs of the following, see Courant-Hilbert, Vol. II, Chapter I, ¶ 3.

$$\frac{du}{ds} = \frac{\partial u}{\partial x}\frac{dx}{ds} + \frac{\partial u}{\partial y}\frac{dy}{ds} = \frac{dx}{ds} + q\frac{dy}{ds}$$

together with (29) to determine the remaining initial values $p(s)$ and $q(s)$, the integral surface being given then by use of (30) and (31). Here as before the characteristics represent intersections of integral surfaces.

The general partial differential equation in n independent variables

$$F(x_1,\dots x_n;\ u;\ p_1,\dots p_n) = 0 \tag{32}$$

is somewhat more general than equation (26) with which we are concerned, since the dependent variable, I, is absent in (26), but this difference is unessential for, if instead of looking for a solution of (32) having the form $u(x_1,\dots x_n)$ we consider it to be given implicitly as

$$\varphi(u;\ x_1,\dots x_n) = 0 ,$$

we can rewrite (32) as a differential equation in the n+1 independent variables $u, x_1,\dots x_n$. We have

$$\sum_{i=1}^{n} \varphi_{x_i} dx_i + \varphi_u du = 0$$

and

$$du = \sum_{i=1}^{n} u_{x_i} dx_i = \sum_{i=1}^{n} p_i dx_i$$

from which we obtain immediately

$$p_i = -\frac{\varphi_{x_i}}{\varphi_u} ,$$

equation (32) becoming

$$F(x_1,\dots x_n;\ u;\ -\frac{\varphi_{x_1}}{\varphi_u},\dots \frac{\varphi_{x_n}}{\varphi_u}) = 0 , \tag{33}$$

in which the dependent variable φ no longer appears. If we

consider the implicit relation (33) solved for φ_u and replace φ by I, u by t, and x_i by u_i we get exactly

(26) $$I_t + L(t: u_1, \ldots u_n: I_{u_1}, \ldots I_{u_n}) = 0 .$$

The characteristic differential equations for (32) are

(34) $$\frac{dx_i}{ds} = F_{p_i} , \quad \frac{du}{ds} = \sum_{i=1}^{n} p_i F_{p_i} , \quad \frac{dp_i}{ds} = -(F_{x_i} + F_u p_i)$$

where we have 2n+1 differential equations in the 2n+1 variable $x_i(s)$, $u(s)$, $p_i(s)$. The corresponding characteristic differential equations for (26) take the form

(35) $$\begin{cases} \dfrac{du_i}{dt} = L_{v_i} , & \dfrac{dv_i}{dt} = - L_{u_i} \\ \dfrac{dI}{dt} = \displaystyle\sum_{i=1}^{n} v_i L_{v_i} - L , & \dfrac{dI_t}{dt} = - L_t \end{cases}$$

where we have set $v_i = I_{u_i}$ and taken t for the parameter along the characteristic curves. Here there are 2n+2 differential equations for the 2n+2 variables $u_i(t)$, $I(t)$, $I_{u_i}(t) = v_i(t)$, and $I_t(t)$. However, we see that since I and I_t do not appear in L, the 2n equations

(18) $$\frac{du_i}{dt} = L_{v_i} \qquad \frac{dv_i}{dt} = - L_{u_i}$$

can be treated as a system in itself, independent of the other two. But these are exactly the equations for the extremals of our variational problem. We thus see that the extremals of the integrals (14) or (15) are given by the projection in the (u,t) space of the characteristics of the corresponding Hamilton-Jacobi partial differential equation (which are in general curves in the (I,u,t) space). In the example

(12) $$I = \int_{\tau}^{t} \sqrt{1 + (\frac{du}{ds})^2}\, ds ,$$

the Hamilton-Jacobi equation is

$$I_t^2 + I_u^2 = 1 ,$$

of which the characteristics are straight lines in the t,u,I space making angles of 45° with the t,u plane, and the projections (extremals) are the previously found straight lines in the t,u plane.

We now come to our chief result that from a complete solution of the Hamilton-Jacobi equation we can construct all the extremals of our variational problem. By a complete solution is meant a solution

$$(36)\qquad I(t;\ u_1,\dots u_n;\ a_1,\dots a_n) - a \qquad |I_{u_i a_j}| \neq 0$$

depending on the n+1 parameters a_i, a (the n+1 st parameter a is additive since only derivatives of I appear in the differential equation). The envelope of the n parameter family resulting from setting $a = f(a_1,\dots a_n)$ is obtained from

$$(37)\qquad I_{a_i}(t;\ u_1,\dots u_n;\ a_1,\dots a_n) - f_{a_i}(a_1,\dots a_n) = 0$$

$$(i = 1,\dots n)\ ,$$

by eliminating the parameters a_i in (36) and (37). For each value of the the intersection of (36) and (37) is a characteristic, (the integral surface given by the envelope is traced out by allowing the a_i to take all their values) so the projection of this characteristic in the t,u space namely (37), is the equation of an extremal (represented in (37) as the intersection of n surfaces). Since the function f is arbitrary the quantities $f_{a_i} = b_i$ can be given arbitrary values, and since $|I_{a_i u_j}| \neq 0$ the n equations (37) can be solved for u_i giving

$$(38)\qquad u_i = u_i(t;\ a_1,\dots a_n;\ b_1,\dots b_n)$$

which is the required 2n parameter family of extremals. The theory of characteristics has been used only for motivation, and we will now prove this statement independently.

If from a complete solution $I(t;\ u_1,\dots u_n;\ a_1,\dots a_n)$, $|I_{u_i a_j}| \neq 0$, of the

(28) $$I_t + L(t,u,I_u) = 0 \ ,$$

we define the functions $u_i(t;\ a_1,\dots a_n;\ b_1,\dots b_n)$ implicitly by

(39) $$I_{a_i}(t;\ u_1,\dots u_n;\ a_1,\dots a_n) = b_i, \qquad i = 1,\dots n$$

and the conjugate functions $v_i(t;\ a_1,\dots a_n;\ b_1,\dots b_n)$ by

(40) $$v_i = I_{u_i}(t;\ u_1,\dots u_n;\ a_1,\dots a_n) \ ,$$

the u_i here to be replaced by their values from (39), we get a 2n parameter family of extremals satisfying the canonical equations $du_i/dt = L_{v_i}$, $dv_i/dt = - L_{u_i}$.

Considering I as a function of t, u_i, and a_i, with u_i and v_i functions of t, a_i, and b_i we differentiate (28) with respect to a_i (t and u_i constant) and (39) with respect to t (a_i and b_i constant) obtaining

$$I_{ta_i} + \sum_{j=1}^{n} L_{v_j} I_{u_j a_i} = 0$$

$$I_{a_i t} + \sum_{j=1}^{n} I_{a_i u_j} \frac{\partial u_j}{\partial t} = 0$$

Subtracting and remembering that $|I_{a_i u_j}| \neq 0$, we have $du_j/dt = L_{v_j}$, the total derivative du_j/dt implying that a_i and b_i are held constant. Similarly differentiating (40) with respect to t and (28) with respect to u_i we have

$$\frac{dv_i}{dt} = I_{u_i t} + \sum_{j=1}^{n} I_{u_i u_j} \frac{du_j}{dt}$$

$$I_{tu_i} + L_{u_i} + \sum_{j=1}^{n} L_{v_j} I_{u_j u_i} = 0 \ .$$

Subtracting we obtain

$$\frac{dv_i}{dt} - L_{u_i} = \sum_{j=1}^{n} I_{u_j u_i}\left(\frac{du_j}{dt} - L_{v_j}\right) = 0 \ ,$$

which completes the proof.

We will now proceed to show that while the Hamilton-Jacobi equation was derived by considering distance measured from a point in the (t,u) space, an integral of the equation in general represents distance measured from a surface in the (t,u) space. This becomes clear geometrically if we take for our complete solution (36) the integral $I(t,u,\tau,k)$ representing distance from the point (τ,k), the (τ,k) being $n+1$ parameters. An envelope I' is constructed by assuming some relation $f(\tau;\ k_1,\ldots k_n) = 0$, and since $I = 0$ for every point (τ,k) satisfying $f = 0$, also $I' = 0$ on the surface $f = 0$. Further, the envelope of "spheres" $I = C$ is the surface $I' = C$, so that the latter locus represents those points (t,u) which are at a distance C from the initial surface $f(\tau,k) = 0$.

To make this more precise we define the distance from a given point (t,q) to a surface $T(\tau,k) = 0$, $T_\tau \neq 0$, to be the minimum distance measured along all extremals through the point which intersect the surface. This is essentially the problem of the free boundary, one endpoint (t,q) being fixed and the other lying on the surface $T(\tau,k) = 0$. The condition that the distance I from the point (t,q) to an arbitrary point (τ,k) on $T = 0$ be stationary with respect to variations of (τ,k) is given by

$$(41) \qquad \delta I = L(\tau,k,\lambda)\delta\tau - \sum_{i=1}^{n} \lambda_i \delta k_i = 0$$

using (24), while we have

$$(42) \qquad T_\tau \delta\tau + \sum_{i=1}^{n} T_{k_i}\delta k_i = 0$$

since (τ,k) is constrained to lie on $T = 0$. Eliminating $\delta\tau$ between (41) and (42) (using $T_\tau \neq 0$) we have

$$\delta I = - \sum_{i=1}^{n} \left(\frac{L(\tau,k,\lambda)}{T_\tau} T_{k_i} - \lambda_i\right)\delta k_i = 0 \ ,$$

and since the δk_i are independent,

$$(43) \qquad \frac{\lambda_i(\tau,k)}{T_{k_i}(\tau,k)} = \frac{L(\tau,k,\lambda)}{T_\tau(\tau,k)} \quad .$$

The n transversality conditions (43) in general serve to select one or more extremals from the point (t,q) to the surface T = 0, and express a relation connecting the tangent plane to the surface (through the T_{k_i}) and the slope of the extremal where it intersects the surface (through the λ_i, which are equivalent to the k_i'). Equation (43) is a direct generalization of equation (62) on p. 31 of Chapter I, and in the case of geodesics where I is actually distance, reduces to orthogonality. Dispensing with the external point (t,q), the n conditions (43) in general serve to determine a unique extremal from the n parameter family through each point on the surface T = 0, thereby forming a field of extremals transverse to the surface T = 0,* at least in some neighborhood of the (t,u) space surrounding a region of the surface. In such a neighborhood we can define a single valued function $\bar{I}(t,u)$ taking the value zero on T = 0. If we consider a curve $t = t(\varepsilon)$, $u_i = u_i(\varepsilon)$ lying in this field, we have an extremal intersecting the curve at each point, with initial values $\tau(\varepsilon)$, $k_i(\varepsilon)$ lying on T = 0. The corresponding distance function $I(t,u,\tau,k)$ as a function of 2n+2 variables satisfies equation (24). The variation δI (which is equal to $\delta\bar{I}$ for the particular variation we have chosen) can be considered the sum of two variations, one with (t,u) fixed, and the other with (τ,k) fixed. The former vanishes in virtue of the transversality condition (41), so that we have

$$(44) \qquad \delta\bar{I} = \sum_{i=1}^{n} l_i \delta u_i - L(t,u,l)\delta t$$

from which follows $\bar{I}_t = -L(t,u,l)$ and $\bar{I}_{u_i} = l_i$, and by

---

*Since only derivatives of T appear in (43), transversality for a surface T = 0 implies transversality for the family of surfaces T = constant.

elimination of I_i, $\bar{I}_t + L(t,u,\bar{I}_u) = 0$, so that the function $\bar{I}$ which we have constructed as the distance from an arbitrary surface $I(\tau,k) = 0$ satisfies the Hamilton-Jacobi equation.

The converse fact that any solution of the Hamilton-Jacobi equation represents the distance from a surface (or, as a degeneration, from a point) is readily verified. From a given integral $I(s,u)$ we construct the n ordinary differential equations

$$\frac{du_i}{ds} = L_{v_i}(s,u_i,v_i) \tag{45}$$

where we have used the notation

$$v_i = I_{u_i}(s,u) \ . \tag{46}$$

The solution of (45) subject to the initial conditions $u_i = q_i$ at $s = t$ are

$$\begin{cases} u_i = u_i(s;\ q_1,\ldots q_n) \ , \\ v_i = v_i(s;\ q_1,\ldots q_n) \ , \end{cases} \tag{47}$$

where the values of u_i have been substituted in (46) to get v_i. Differentiating (46) with respect to s, q being constant

$$\frac{dv_i}{ds} = I_{u_i s} + \sum_{j=1}^{n} I_{u_i u_j} \frac{du_j}{ds}$$

$$= I_{u_i s} + \sum_{j=1}^{n} I_{u_i u_j} L_{v_j} \ .$$

Differentiating $I_s + L(s,u,v) = 0$ with respect to u_i we have

$$I_{su_i} + L_{u_i} + \sum_{j=1}^{n} L_{v_j} \frac{\partial v_j}{\partial u_i} = I_{su_i} + L_{u_i} + \sum_{j=1}^{n} L_{v_j} I_{u_j u_i} = 0 \ ,$$

and subtracting from the previous result

$$\frac{dv_i}{ds} - L_{u_i} = \sum_{j=1}^{n} I_{u_i u_j}\left(\frac{du_j}{ds} - L_{v_j}\right) = 0 \ .$$

The constructed functions u and v are therefore an n parameter family of extremals. Furthermore, they are transversal to the surface I = const. For from

$$I_s + L(s,u,v) = 0 \quad \text{and} \quad v_i = I_{u_i}(s,u)$$

we have immediately

$$\frac{v_i(s,u)}{I_{u_i}(s,u)} = \frac{L(s,u,v)}{I_s(s,u)}$$

which are exactly the transversality conditions (43) for the surface $I(s,u) = \text{const}$. It remains to be shown that I is actually the distance function along this family of transverse extremals. Taking $I(\tau,k) = 0$, we have

$$I = \int_\tau^t \frac{dI}{ds} = \int_\tau^t (I_s + \sum_{i=1}^{n} I_{u_i} \frac{du_i}{ds})ds$$

$$\int_\tau^t (\sum_{i=1}^{n} v_i u_i' - L(s,u,v))ds$$

which is exactly (17).

The foregoing concepts can be interpreted in terms of the propagation of light through a medium having a variable index of refraction. The light rays (extremals) are given as paths of least time (I is a minimum). The construction of solutions as envelopes is exactly Huyghen's Principle for the construction of wave fronts.

4. The two body problem. We consider the problem of determining the motion of two bodies of mass m_1 and m_2 acted on only by the Newtonian gravitational force between them,

$$F = \frac{Gm_1m_2}{(r_1 + r_2)^2} ,$$

where r_1 and r_2 are the distances of m_1 and m_2 from the center of mass C, which we may consider fixed. From the relation $m_1r_1 = m_2r_2$, we have

$$F = -\frac{Gm_1m_2}{r_1^2} \quad \text{where } m_2' = \frac{m_2^3}{(m_1 + m_2)^2}$$

that the problem is reduced to that of a fixed mass m_2' attracting mass m_1 at a distance r_1. With C as origin we take coordinates x nd y in the plane determined by the initial position and velocity m_1. The motion is described as making the integral $(T-V)dt$ ationary, where T and V are the kinetic and potential energies espectively. Taking $m_1 = 1$ we have

$$T = (x^2 + y^2)/2$$

$$V = -k^2/\sqrt{x^2 + y^2}$$

$$F = T - V = (x^2 + y^2)/2 + k^2/\sqrt{x^2 + y^2} \ .$$

educing the canonical form we use the notation $x = p, \quad y = q$ nd have $F_x = p$, $F_y = q$ yielding

$$L(t,x,y,p,q,) = \frac{1}{2}(p^2 + q^2) - k^2/\sqrt{x^2 + y^2}$$

or the Hamilton function and

$$\varphi_t + \frac{1}{2}(\varphi_x^2 + \varphi_y^2) - k^2/\sqrt{x^2 + y^2} = 0$$

or the Hamilton-Jacobi equation. Changing to polar coordinates, $= r \cos \theta$, $y = r \sin \theta$, this reduces to

$$\varphi_t + \frac{1}{2}(\varphi_r^2 + \frac{1}{r^2}\varphi_\theta^2) = \frac{k^2}{r}.$$

two parameter solution can be found by writing as the sum of hree functions of t, r, and θ respectively, in particular

$$\varphi = \alpha t + \beta\theta + R(r) \ .$$

ubstituting in the differential equation we solve for R obtaining

$$\varphi = \alpha t + \beta\theta + \int_{r_0}^{r} \sqrt{\frac{2k^2}{\rho} - \frac{\beta^2}{\rho^2} - 2\alpha} \ d\rho \ .$$

o solve for the extremals we differentiate with respect to the arameters obtaining

$$t - \int_{r_o}^{r} \frac{d\rho}{\sqrt{\frac{2k^2}{\rho} - \frac{\beta^2}{\rho^2} - 2\alpha}} = t_o$$

$$\theta - \beta \int_{r_o}^{r} \frac{d\rho}{\rho^2\sqrt{\frac{2k^2}{\rho} - \frac{\beta^2}{\rho^2} - 2\alpha}} = \theta_o \ ;$$

t_o and θ being arbitrary constants. The second equation gives θ as function of r, i.e. the particle path, while the first gives r as a function of time. The second equation can be integrated using the substitution $\rho = 1/\sigma$ giving

$$\theta = \theta_o - \text{arc sin} \frac{\frac{\beta^2}{k^2 r} - 1}{\sqrt{1 + \frac{2\alpha\beta^2}{k^4}}}$$

Writing $\delta = \frac{\beta^2}{k^2}$ and $\varepsilon^2 = \sqrt{1 + \frac{2\alpha\beta^2}{k^4}}$ we have the conic of eccentricity ε

$$r = \frac{\delta}{1 - \varepsilon^2 \sin(\theta - \theta_o)} .$$

5. The Homogeneous Case-- Geodesics. Up to now it has been assumed that $|F_{u_i' u_j'}| \neq 0$, but this condition excludes the important case where F is homogeneous of first order in u_i'. For this case we have $|F_{u_i' u_j'}| \equiv 0$, since if F is homgeneous of first order, $F_{u_i'}$ is homogeneous of zeroth order, and applying Euler's homogeneity condition

$$\sum_{j=1}^{n} u_j' F_{u_i' u_j'} = 0 ,$$

we see the determinant $|F_{u_i' u_j'}| \equiv 0$. If in addition to being homogeneous of first order in u_i', F is independent of s (which is the case for variational problems in parametric form), the problem can be reduced to the form already discussed by taking u_n for the variable of integration s. We then have

(48) $$J = \int_{\tau}^{t} F\left(u_1,\ldots u_n;\ \frac{du_1}{du_n},\ldots\frac{du_{n-1}}{du_n}\right)du_n ,$$

as our integral, and the Hamilton-Jacobi equation becomes

(49) $$J_{u_n} + L(u_1,\ldots u_{n-1};\ J_{u_1},\ldots J_{u_{n-1}}) = 0$$

where

$$L(u_1,\ldots u_{n-1};\ v_1,\ldots v_{n-1}) = \sum_{i=1}^{n-1} p_i v_i - F(u_1,\ldots u_{n-1};\ p_1,\ldots p_{n-1})$$

with $p_i = \frac{du_i}{du_n}$, $v_i = F_{p_i}$. The equation (49) is nothing more than the homogeneity condition, since from

$$\sum_{i=1}^{n} \bar{p}_i F_{\bar{p}_i} - F = 0 , \qquad \bar{p}_i = \frac{du_i}{d\tau} ,$$

substituting $p_i = \bar{p}_i/\bar{p}_n$ we have

(50) $$F_{\bar{p}_n} + \sum_{i=1}^{n-1} p_i F_{\bar{p}_i} - F(u_1,\ldots u_n;\ p_1,\ldots p_{n-1},1) = 0 .$$

The expressions (25) for the derivatives J_t, J_{u_i} are true even with $|F_{u_i u_j}| \equiv 0$,* and if we substitute J_{u_i} for $F_{\bar{p}_i} = v_i$ (i.e. l_i) in (50) we get exactly (49).

Let us illustrate another method of attack with the case of geodesics on an n dimensional manifold. We have

(51) $$I = \int_{\tau}^{t} \sqrt{Q}\, ds ,$$

where $Q = \sum_{i,k=1}^{n} g_{ik} u_i' u_k'$, the g_{ik} being functions of the u_i.

---

*This condition does not prevent the use of the Legendre transformation, but only invalidates its inversion; in other words the extremals of (9) are included in those of (17), and this fact is all that is needed in deriving (25).

From (25) we have

$$(52)\qquad \begin{cases} I_t = 0 \\ I_{q_i} = F_{u'_i}\big|_{s=t} = \dfrac{1}{\sqrt{Q}} \displaystyle\sum_{k=1}^{n} g_{ik} q'_k \ . \end{cases}$$

If we denote by (g^{ik}) the matrix reciprocal to (g_{ik}) (i.e. $\sum_{j=1}^{n} g_{ij} g^{jk} = \delta_i^k$), solving (52) for q'_k gives

$$(53)\qquad \frac{1}{\sqrt{Q}}\, q'_k = \sum_{i=1}^{n} g^{ik} I_{q_i} \ .$$

From the homogeneity relation

$$\sum_{k=1}^{n} q'_k F_{u'_k} = \sum_{k=1}^{n} q'_k I_{q_k} = \sqrt{Q} \ ,$$

so that multiplying through (53) by I_{q_k} and summing with respect to k we obtain

$$(54)\qquad \sum_{i,k=1}^{n} g^{ik} I_{q_i} I_{q_k} = 1$$

which is a partial differential equation in the I_{q_i} taking the place of the Hamilton-Jacobi equation. The connection between (54) and the Hamilton-Jacobi equation is clarified by a consideration of the Euler equations of (51). We have

$$\frac{d}{ds}\left(\frac{\partial \sqrt{Q}}{\partial u'_i}\right) - \frac{\partial \sqrt{Q}}{\partial u_i} = 0$$

or

$$(55)\qquad \frac{d}{ds}\left(\frac{\partial Q}{\partial u'_i} \cdot \frac{1}{\sqrt{Q}}\right) - \frac{\partial Q}{\partial u_i} \cdot \frac{1}{\sqrt{Q}} = 0 \ .$$

If for any admitted curve (and in particular for the extremals) we take the parameter s to be proportional to the

are length then Q is a constant, and (55) reduces to

$$\frac{d}{ds}\left(\frac{\partial Q}{\partial u_i'}\right) - \frac{\partial Q}{\partial u_i} = 0 \tag{56}$$

for the function $Q(u_1, \ldots u_n : u_1', \ldots u_n') = C$. Now, (56) suggests that we look for extremals of the integral

$$J = \int_{t_0}^{t} Q\,ds \tag{57}$$

subject to the condition $Q = C$, where Q is homogeneous but of second order, so the Hamilton-Jacobi theory can be applied directly. We have

$$Q = \sum_{i,k=1}^{n} g_{ik} p_i p_k \quad , \quad p_i = u_i'$$

$$v_i = Q_{p_i} = 2 \sum_{k=1}^{n} g_{ik} p_k \; .$$

Solving for p_k

$$p_k = \frac{1}{2} \sum_{i=1}^{n} g^{ik} v_i \; ,$$

$$H = \sum_{i=1}^{n} p_i v_i - Q$$

$$= Q$$

$$= \sum_{i,k=1}^{n} g_{ik} p_i p_k$$

$$= \frac{1}{4} \sum_{i,k=1}^{n} g^{ik} v_i v_k \tag{58}$$

and the differential equation is

(59) $$J_t + \frac{1}{4} \sum_{i,k=1}^{n} g^{ik} J_{u_i} J_{u_k} = 0 .$$

In order to find an integral of (59) containing n parameters we try a solution in the form $J(t; u_1,\ldots u_n) = f(t) + J^*(u_1,\ldots u_n)$ and it follows immediately that $f'(t)$ must be independent of t, i.e. $f(t) = \alpha t$. J^* then satisfies the equation

$$\alpha + \frac{1}{4} \sum_{i,k=1}^{n} g^{ik} J_{u_i} J_{u_k} = 0 .$$

This differential equation in general allows an n parameter solution, so α may be given the specific value $- 1/4$, and we have

(60) $$\sum_{i,k=1}^{n} g^{ik} J_{u_i} J_{u_k} = 1 ,$$

which is (54) again.

From (58) it is apparent that Q is constant (viz. $Q = 1/4$) for any solution of (60), so that any n parameter integral of (60) can be used to obtain the extremals of (51) as well as those of (57). Using the methods outlined here it is possible to obtain the geodesics on an ellipsoid. The surfaces

(61) $$\frac{x^2}{a^2+s} + \frac{y^2}{b^2+s} + \frac{z^2}{c^2+s} = 1$$

in the parameter s, where $a > b > c$ are given constants, represent confocal ellipsoids, one-sheeted hyperboloids, and two-sheeted hyperboloids as $-c^2 < s$, $-b^2 < s < -c^2$, $-a^2 < s < -b^2$ respectively. At a given point (x,y,z) (61) is a cubic in s with three real roots (s_1, s_2, s_3) representing three mutually orthogonal surfaces one of each type, through the point (x,y,z).* The parameters (s_1, s_2, s_3) can be taken as a new coordinate system, and in particular,

details see Courant-Hilbert, Vol. I, p. 195.

taking our given ellipsoid as $s_1 = 0$, the remaining two coordinates can be used as parameters on the ellipse, with the transformation formulas (setting $-s_1 = u$, $-s_2 = v$)

$$x = \sqrt{\frac{a(a-u)(a-v)}{(a-b)(a-c)}}$$

$$y = \sqrt{\frac{b(b-u)(v-b)}{(b-c)(a-b)}}$$

$$z = \sqrt{\frac{c(u-c)(v-c)}{(a-c)(b-c)}} \; .$$

The values of

$$g_{11} = (\frac{\partial x}{\partial u})^2 + (\frac{\partial y}{\partial u})^2 + (\frac{\partial z}{\partial u})^2,$$

$$g_{12} = g_{21} = (\frac{\partial x}{\partial u})(\frac{\partial x}{\partial v}) + (\frac{\partial y}{\partial u})(\frac{\partial y}{\partial v}) + (\frac{\partial z}{\partial u})(\frac{\partial z}{\partial v}) \; ,$$

$$g_{22} = (\frac{\partial x}{\partial v})^2 + (\frac{\partial y}{\partial v})^2 + (\frac{\partial z}{\partial v})^2,$$

become

$$g_{11} = (u-v)A(u)$$

$$g_{12} = 0$$

$$g_{22} = (v-u)A(v) \; ,$$

where

$$A(w) = \frac{1}{4}\,\frac{w}{(a-w)(b-w)(c-w)} \; .$$

Also

$$g^{11} = \frac{1}{(u-v)A(u)}$$

$$g^{12} = 0$$

$$g^{22} = \frac{1}{(v-u)A(v)} \; .$$

Equation (60) takes the form

$$A(v)J_u^2 - A(u)J_v^2 = (u-v)A(u)A(v) \quad . \tag{62}$$

To get a one* parameter solution of (62) we try $J(u,v) = \Phi(u) + \Psi(v)$, and separation of variables gives

$$\frac{\Phi'(u)^2}{A(u)} - u = \frac{\Psi'(v)^2}{A(v)} - v = \alpha$$

so that

$$J(u,v,\alpha) = \int \sqrt{(u + \alpha)A(u)}\, du + \int \sqrt{(v + \alpha)A(v)}\, dv \quad .$$

Differentiating with respect to α we get the two parameter family of geodesics

$$\int \sqrt{\frac{A(u)}{u + \alpha}}\, du + \int \sqrt{\frac{A(v)}{v + \alpha}}\, dv = \beta \quad . \tag{63}$$

This can be solved for u or v in terms of elliptic functions

6. Sufficient Conditions. Analogous to the vanishing of the first derivative of a function f(x) of a single variable we have found the vanishing of the first variation (leading to the Euler equations) as a necessary condition for the minimum of a variational problem. Corresponding to the sufficient condition $f''(x) > 0$ we might look for sufficient conditions in a variational problem by investigating the second variation. Although such considerations can lead to further necessary conditions (e.g. the Legendre condition $F_{u'u'} \geq 0$, p. I-53) they can never lead to a sufficient condition. The reason for this is that in order to derive a sufficient condition we must consider all possible admissible variations, i.e. $\Phi(x,\varepsilon) = u(x) + \zeta(x,\varepsilon)$ where $\zeta(x,\varepsilon)$ is an arbitrary admissible function with zero boundary values, vanishing for $\varepsilon = 0$. However, it is to construct an admissible variation (e.g. $\zeta = (1-x)\varepsilon \sin x/\varepsilon^2$ $0 \leq x \leq 1$) for which $\zeta_x(x,\varepsilon)$ does not approach zero for $\varepsilon \longrightarrow 0$. In this case the varied integral

*Because of the homogeneity this problem has essentially one dependent variable. This is obvious if we take v, say, for the parameter t.

$$I(\varepsilon) = \int_{x_0}^{x_1} F(x, \varphi(x,\varepsilon), \varphi_x(x,\varepsilon))dx$$

does not converge to the desired integral

$$I = \int_{x_0}^{x_1} F(x,u,u')dx$$

as $\varepsilon \to 0$, and the variational problem does not reduce to that of minimizing a function of the single variable ε. A variation $\zeta(x,\varepsilon)$ which satisfies both conditions $\zeta(x,\varepsilon) \to 0$, $\zeta_x(x,\varepsilon) \to 0$ as $\varepsilon \to 0$ is called a weak variation, and geometrically means that the curve $u(x)$ is compared with curves that approximate $u(x)$ in slope as well as position. A curve $u(x)$ which minimizes an integral with respect to all weak variations is called a weak relative minimum (relative referring to the fact that $u(x)$ is a minimum only with respect to curves in its neighborhood). For example, consider

$$I = \int_0^1 [u'(x)^2 - u'(x)^3]dx$$

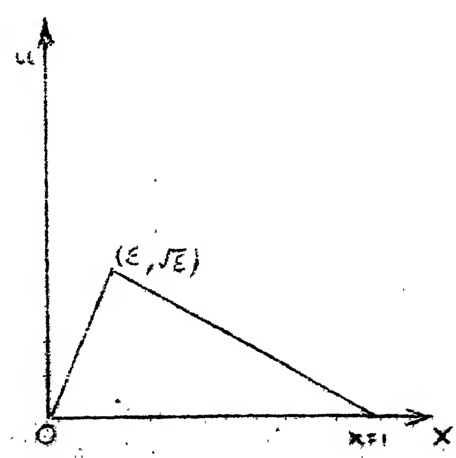
subject to $u(0) = u(1) = 0$. The extremals are straight lines, and there is a unique extremal $u = 0$ satisfying the boundary conditions. The value of I is zero for $u = 0$, and is positive for all curves satisfying the condition $u'(x) < 1$, so $u = 0$ minimizes I with respect to this restricted class of neighboring curves.

However, by taking the admissible curve

$$u = \frac{x}{\sqrt{\varepsilon}} \qquad (0 \le x \le \varepsilon)$$

$$u = \frac{\sqrt{\varepsilon}(1-x)}{1-\varepsilon} \qquad (\varepsilon \le x \le 1)$$

which approaches the extremal $u = 0$ uniformly we can make I negative.



We shall now proceed to establish Weierstrass' sufficient condition for an extremal to be a strong relative minimum for the integral

$$(64) \qquad I(u) = \int_{\tau}^{t} F(s,u,u')ds$$

subject to the boundary conditions

$$A: \quad (s = \tau,\ u = k) \qquad B: \quad (s = t,\ u = q)\ .$$

In order to do this we must compare the values of the integral (64) over the extremal between A and B, (I_e), and an arbitrary curve C in its neighborhood, (I_c). By expressing I_e as an integral along the path C, we shall reduce this problem to a comparison of the integrands alone. Assuming that the extremal in question can be imbedded in a field of extremals*, we define a slope function $p(u,s)$ as the slope of the extremal through the point $P = (u,s)$. Further we define the single valued distance function $\bar{I}(P)$ (see p. II-26)** which has the property that $I_e = \bar{I}(B) - \bar{I}(A)$. The differential $d\bar{I} = \bar{I}_s ds + \bar{I}_u du$ is exact and we have

$$(65) \qquad I_e = \int_A^B (\bar{I}_s ds + \bar{I}_u du) = \int_{\tau}^{t} (\bar{I}_s + u'\bar{I}_u)ds$$

where the integration is taken over any curve joining A to B, and in particular over the curve C. Using (25) and (16) we have $\bar{I}_s = F(s,u,p) - pF_p(s,u,p)$ and $\bar{I}_u = F_p(s,u,p)$ so that (65) reduces to

$$I_e = \int_{\tau}^{t} [F(s,u,p) + (u' - p)F_p(s,u,p)]ds$$

and finally we have

$$(66) \quad \Delta I = I_c - I_e = \int_{\tau}^{t} [F(s,u,u') - F(s,u,p) - (u'-p)F_p(s,u,p)]ds$$

---

*This condition will be investigated in section 7.

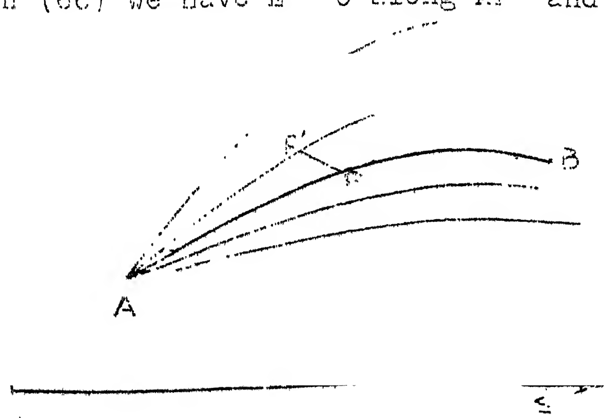
**For the existence of $\bar{I}$ we must in general have a field of extremals transverse to some surface, but for the case of only one dependent variable u this condition is automatically satisfied for any field.

where the integration is taken along the path C, u' representing the slope of the curve C and p the slope of the field at each point. The integrand of (66)

(67) $$- F(s,u,p) - (u'-p)F_p(s,u,$$

is known as Weierstrass' E function. If the condition $E \geq 0$ is satisfied at each point s, u, p in the field for all values of u', then from (66) $\Delta I \geq 0$ for all curves C in the field; in other words $I_c \geq I_e$ and the extremal joining A and B is actually a minimum. We therefore conclude that a sufficient condition for an extremal joining two points to be a minimum is that it be possible to imbed it in a field throughout which the condition $E(s,u,p,u') \geq 0$ holds for all values of u'. Obviously if $E > 0$ for all $u' \neq p$, then we have a proper minimum with $I_c > I_e$ for C not the extremal in question.

We can show that the weaker condition $E \geq 0$ along the extremal for all values of u' is <u>necessary</u> for a minimum. For, let us assume that the extremal between A and B is a minimum and at a point P on it $E < 0$ for some value u'. The extremal can be imbedded in a field emerging from A (that this condition is necessary will be shown in the next section). By continuity $E < 0$ on a small line segment PP' of slope u' where P' can be joined to A by an extremal of the field. Now taking the path AP'PB for the curve C in (66) we have $E = 0$ along AP' and PB, with $E < 0$ on P'P, thereby reaching the contradiction $I_c < I_e$. Since P was any point and u' any value, we conclude that $E \geq 0$ along an extremal for all u' is a necessary condition for a minimum.



If from $E > 0$ for all u' along an extremal we could conclude the sufficient condition that $E \geq 0$ in a neighborhood for all u' then the analogy with functions of a single variable

would be complete (i.e. $f''(\lambda) \ge 0$ necessary and $f''(\lambda) > 0$ sufficient for a minimum). Unfortunately this inference is not true unless we impose the restriction that u' be uniformly bounded in the neighborhood (see Bolza, Lectures on the Calculus of Variations, p. 99).

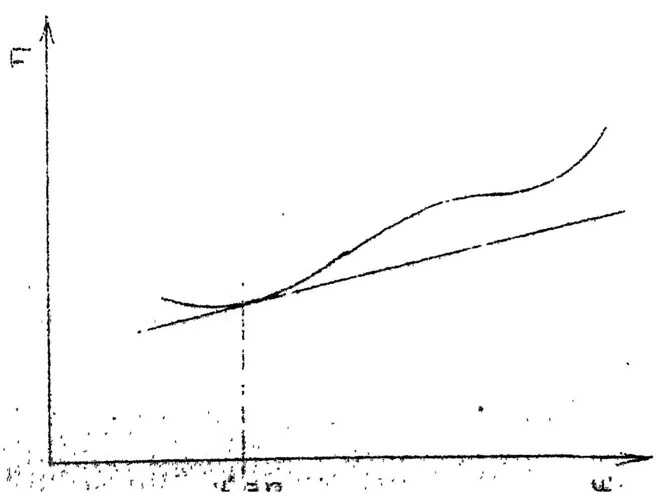
The significance of the Legendre condition can be appreciated by comparison with the E function. Using the theorem of the mean we have

$$F(s,u,u') = F(s,u,p)+(u'-p)F_p(s,u,p)+\frac{1}{2}(u'-\bar{p})^2F_{pp}(s,u,\bar{p})$$

where $\bar{p} = p + \theta(u'-p)$, $0 < \theta < 1$, so (67) becomes

(68) $$E(s,u,p,u') = \frac{1}{2}(u' - \bar{p})^2F_{pp}(s,u,\bar{p})$$

$\bar{p}$ being some value between p and u'. From (68) we see that if $F_{u'u'}(s,u,u') \ge 0$ at all points of the field for arbitrary values of u', then also $E \ge 0$, so this condition is sufficient. A problem for which $F_{u'u'}(s,u,u') > 0$ for all values of the quantities s, u, u' is said to be regular, and for such a problem the existence of a field of extremals guarantees a proper minimum. The connection between the Legendre condition and Weierstrass' E function may be interpreted geometrically in the following way. Considering $F(s,u,u')$ as a function of the direction u' at a fixed point of the field (s,u) (thereby fixing $p(s,u)$), $E(s,u,p,u') = 0$ is the equation of the tangent to F at the point $u' = p$, and $E \ge 0$ for all u' states that $F(s,u,u')$ lies above the tangent line. The condition $F_{u'u'}(s,u,u') \ge 0$ for



all u' means that the curve is convex, and therefore lies above the tangent line. If this is true for all points (s,u), then the same statement can be made of E. The necessary condition $E \geq 0$ along an extremal includes the weak Legendre $F_{pp}(s,u,p) \geq 0$, since $E \geq 0$ for all u' implies convexity at the tangent point u' = p. Although the strong Legendre condition $F_{pp}(s,u,p) > 0$ is neither necessary nor sufficient for a strong minimum it is sufficient for a weak minimum. This is clear since from $F_{pp}(s,u,p) > 0$ on the extremal we can conclude that $F_{u'u'}(s,u,u') > 0$ for (s,u) in some neighborhood of the extremal and u' in the neighborhood of p. This means, however, that $E \geq 0$ for weak variations (by (68)) which is sufficient for a minimum. Although $F_{u'u'}(s,u,u') > 0$ along an extremal for all u' is somewhat stronger than $E > 0$ along an extremal, it is still not sufficient for a strong minimum.

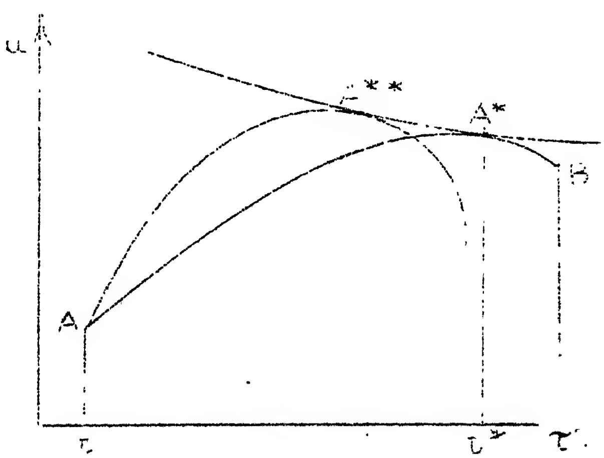
7. Construction of a Field - The Conjugate Point. We have observed that an essential point in the theory of sufficient conditions is the possibility of imbedding a given extremal in a field. We shall now see that this can always be done if the endpoints of the extremal are not too far apart. In general a one parameter family of extremals through a point will constitute a field up to its envelope.

But first of all it is important to note that if an extremal can be imbedded in a field H of extremals so that $E(s,u,p,u') \geq 0$ over the field, then the condition $E(s,u,p^*,u') \geq 0$ will also hold for any other field H^* over their common (s,u) domain. For, supposing that $E < 0$ for some u' at a point (s,u,p^*), we can construct a curve C as on p. 31 for which $I_c < I_e$. However, this is impossible since the curve C lies in the field H in which I_e is proved to be a minimum. We therefore need consider the construction of but a single imbedding field. Moreover, it can be shown that if any imbedding field exists, then the family of extremals through one endpoint is such a field (see Bolza, p. 57). We therefore consider a family of extremals u(s,a) through the end-

point A: (t,q). The first point at which $u(s,a_o)$ is intersected again by a neighboring extremal $u(s, a_o+\varepsilon)$ is given by the simultaneous solution of

$$(69) \qquad \begin{cases} u = u(s,a_o) \\ 0 = u_a(s,a_o) \ . \end{cases}$$

This first point of intersection A^*: (t^*,q^*), is called the conjugate point to A. If an envelope exists then A^* is the intersection of $u(s,a_o)$ with the envelope. Supplementing our previous necessary and sufficient conditions we now have the necessary condition $t^* \leq \tau$ and the sufficient condition $t^* < \tau$ for a minimum. The necessity of $t^* \leq \tau$ may be indicated by the following geometrical argument. If the conjugate point A^* (i.e. intersection with the envelope) lies before B, then taking another extremal with conjugate point A^{**} we can construct the admissible curve $AA^{**}A^*B$, where the arcs AA^{**} and A^*B are extremals and $A^{**}A^*$ is along the envelope. The envelope has the same slope as the field at every point, so $E = 0$ along it. Using (66) we have



$$I_{AA^{**}} + I_{A^{**}A^*} = I_{AA^*} \ .$$

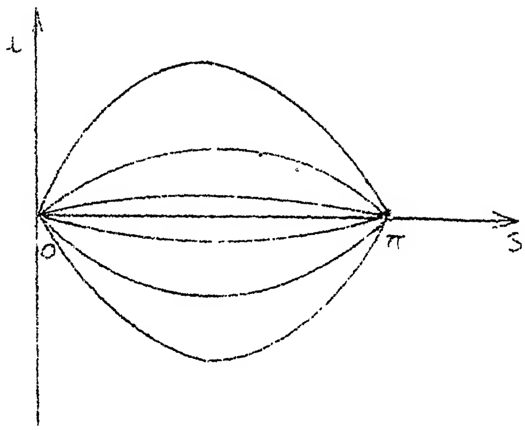
However, since the envelope is not an extremal in general, by connecting A^{**} and A^* by an extremal the value of $I_{AA^{**}A^*}$ can be reduced, so I_{AA^*} is not a minimum. This necessary condition can be established more rigorously by consideration of the second variation (cf. Bolza, p. 57). The sufficiency of $t^* > \tau$ for the existence of a field of extremals can be seen as follows. We have $u_a(s,a_o) \neq 0$ for $t < s \leq \tau < t^*$, and assuming u_a is continuous we may take $u_a > 0$ in this interval. By continuity we also have $u_a(s, a_o+\varepsilon) > 0$ for

$|\varepsilon|$ small enough, so that at a fixed point s, u is a monotonic increasing function of a, covering a neighborhood of $\varepsilon = 0$ once and only once.

Let us take the example

$$J = \int_0^t (u'^2 - u^2)ds \qquad u(0) = 0 .$$

The extremals are $u = a \sin s + b \cos s$, and through the point (0,0) there is the field $u = a \sin s$. We have $F_{u'u'} = 2$, $p = u \cot s$, $E = (u' - p)^2$, and the conjugate point to (0,0) is $(0,\pi)$. The problem is a regular one, so the extremal joining (0,0) to any point (t,u) is a minimum of $t < \pi$.



Summarizing the results of this and the previous section we have as necessary conditions for a strong minimum the Euler equations, the weak Legendre condition $F_{pp}(s,u,p) \geq 0$ which is included in Weierstrass' condition $E(s,u,p,u') \geq 0$ along the extremal, and the conjugate point condition $t^* \geq \tau$. As sufficient conditions we have the Euler equations together with either $F_{u'u'}(s,u,u') \geq 0$ or $E(s,u,p,u') \geq 0$ in a neighborhood, and $t^* > \tau$. For a weak minimum we have as necessary conditions the Euler equations, $F_{pp}(s,u,p) \geq 0$, and $t^* \geq \tau$, which become sufficient on dropping the equality signs.

Problem: Show that for $F = u\sqrt{1 + (u')^2}$ the tangents to an extremal at a point A and at its conjugate point A^* intersect on the s axis.

III. DIRECT METHODS IN THE CALCULUS OF VARIATIONS

Introduction

The so-called direct methods in the calculus of variations represent a relatively modern trend which has established the calculus of variations in a dominating position in mathematical analysis.

Two general points of view in the calculus of variations are relevant for various domains of mathematics, namely the formation of invariants and covariants in function spaces, and the characterization of mathematical entities by extremum properties. We shall concentrate on the second topic. Such a characterization is useful in many fields of mathematics, and often serves to simplify more involved deductions. For example, in the theory of numbers the greatest common divisor of two integers, a and b, can be characterized as the minimum of the expression $|ax + by|$, where for x and y all integers are admitted "to competition". In this course we shall confine ourselves to the field of mathematical analysis.

In the mathematical treatment of physical phenomena it is often expedient to use formulations by means of which the quantities under consideration appear as extrema. An example of that is Fermat's Principle in optics. In mechanics the principle of stable equilibrium has a basic importance: a system is in stable equilibrium if, and only if, the potential energy is a minimum. For elementary mechanics equilibrium conditions are expressed by certain local conditions (vanishing of the sum of all forces and moments). With the help of the calculus of variations it becomes possible to

to characterize a state of equilibrium by one value only, the extremum of a functional. The variational equations furnish, then, the local conditions.

The classical methods of the calculus of variations can be considered as indirect methods, in contrast to the modern direct methods. The distinction is not absolute, for many "modern" ideas appear already at the beginning of the calculus of variations, although without the degree of precision now attained.

Generally speaking, the direct methods aim at solving boundary value problems of differential equations by reducing them to an equivalent extremum problem of the calculus of variations, and then by attacking this problem directly, a procedure which is, so to speak, the reverse of classical calculus of variations.

The most notable example of the direct approach goes back to Gauss and William Thompson. They considered the boundary value problem of the harmonic equation $\Delta u = 0$ for a domain G in the xy-plane, under the condition that the function u be regular in G and attain prescribed continuous boundary values at the boundary.

The classical formalism of the calculus of variations for the integral

$$D[\varphi] = \iint_G (\varphi_x^2 + \varphi_y^2)\,dxdy$$

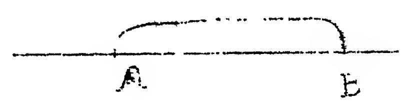
shows:

If $u(x,y)$ furnishes the minimum of the integral when for φ all functions are admitted to competition which are continuous in G and on its boundary, attain the prescribed boundary values, and possess continuous first and second derivatives in G, then $u(x,y)$ is the solution of the boundary value problem for $D[u] = 0$ in G.

Gauss and Thompson thought that, since the integral $D[\varphi]$ is always positive, it must have a minimum; hence the existence of a solution of the boundary value problem appeared established. This reasoning was later resumed by Dirichlet, and a decisive use of it, under the name of Dirichlet's Principle, was made by Bernhard Riemann in his epoch-making investigations on the theory of functions. However, it was soon observed by Weierstrass that the reasoning suffered from a very serious gap. A set of non-negative numbers necessarily has a greatest lower bound, but this lower bound need not be a minimum actually attained in the set. To make Dirichlet's Principle a cogent proof, the existence of a minimum, rather than a greatest lower bound, has to be established. That this is not a trivial matter can be seen from many simple examples of extremum problems apparently "reasonable".

(a) Find the shortest curve from A to B with the condition that it should be perpendicular to AB at A and B.

The length of the admissible curves has a greatest lower bound, namely $\overline{AB}$; however, no shortest curve exists.

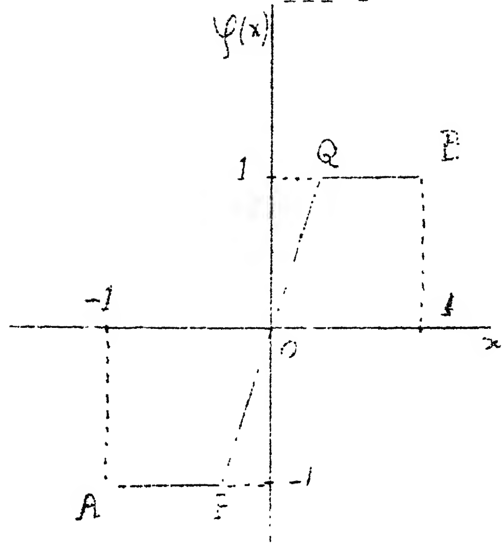


(b) Find a function $\varphi(x)$ continuous, having a piecewise continuous derivative, for which the integral

$$I = \int_{-1}^{1} \varphi'^{2}\,dx$$

attains the smallest possible value, with the boundary conditions:

$$\varphi(-1) = -1\,, \qquad \varphi(1) = 1\,.$$



The integral is always positive and has a greatest lower bound, namely, 0. However, we can make it as small as we want, by taking for $\varphi(x)$ the function represented by APQB. If ε is the distance from P to the point (0,-1),

$$\varphi'(x) = \frac{1}{\varepsilon} \text{ for } -\varepsilon < x < \varepsilon ,$$

$$\varphi'(x) = 0 \text{ for } -1 < x < -\varepsilon \quad \text{or} \quad \varepsilon < x < 1 .$$

Hence

$$I = \int_{-\varepsilon}^{+\varepsilon} \frac{x^2}{\varepsilon^2}\, dx = \frac{2}{3}\, \varepsilon ,$$

and it can be made as small as we want. But the only function for which $I = 0$ is $\varphi(x) = c$, and it does not correspond to a curve through the given end-points.

Weierstrass' criticism was met only much later by Hilbert, when he succeeded in 1900 in establishing the existence of a minimum for problems involving the integral $D[\varphi]$, and thus opened the way for broad developments in the calculus of variations.

The direct methods thus inaugurated marked a great progress in pure and applied analysis, all of it based upon the reduction of boundary value problems to minimum problems. Three related goals are envisaged by such methods:

a) Existence proofs for solutions of boundary value problems,

b) Analysis of the properties of these solutions,

c) Numerical procedures for calculating the solutions.

This last point of view has been stressed by Rayleigh and,

in a broader way, by Walter Ritz, who developed powerful numerical methods of great importance for physics and engineering.

A few examples will show how certain results concerning minimum problems can be attained directly.

By the formal approach the isoperimetric property of the circle (or of the sphere) is reduced to a differential equation, supplemented by certain sufficient conditions, for instance, Weierstrass' conditions. However, the direct approach leads to the result in a straightforward way.

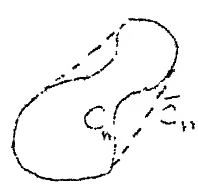
On the basis of Steiner's proof* or of the classical calculus of variations, we may assume as proved that the circle is the solution, provided a solution is assumed to exist. Hence only the existence of a solution needs to be proved.

We consider the following problem: Among all continuous, closed curves C having a given length L, find one which makes the enclosed area $A(C)$ a maximum.

Since any admissible curve C can be completely enclosed in a circle of radius $L/2$, $A(C) \leq \pi L^2/4$; hence a least upper bound M exists for all the areas, and a maximizing sequence $C_1, \ldots, C_n, \ldots$ of admissible curves exists such that

$$A_n(C_n) \rightarrow M \quad \text{for } n \rightarrow \infty .$$

Each C_n can be assumed to be a convex curve, for if not, it could be replaced by a convex admissible curve of larger area: C_n is first replaced by its "convex hull" $\overline{C}_n$; then $\overline{C}_n$ is magnified into a similar admissible curve of length L, $\overline{\overline{C}}_n$, and



$$A(C_n) < A(\overline{C}_n) < A(\overline{\overline{C}}_n).$$

---
* See this proof in Courant-Robbins, What is Mathematics, p.375

(This reasoning is not valid in 3 dimensions. Let us consider a sphere with a long spine. It can be replaced by a convex surface enclosing a larger volume; but this surface will also have a larger area, which prevents us from extending the argument to more than 2 dimensions.)

We now make use of the following theorem: In a sequence of convex curves lying in a closed domain there is a subsequence which converges to a closed convex curve.

Hence a subsequence of curves C_n converges to a convex curve C; since the area of a sequence of convex curves depends continuously on the curves, and the areas A_n of C_n converging to M, $A(C) = M$.

We now make use of a very important fact, the lower semi-continuity of length:

If C_n converges to C, then

$$\underline{\lim}\, L(C_n) \geq L(C) .$$

In the present case, we then have

$$L(C) \leq L .$$

The equality sign, however, must hold, since, if $L(C) < L$, C could be modified into a curve of length L, whose area would then exceed M. Thus the existence of a curve of maximum area and length L is established.

Lower semi-continuity of length is only an example of a property which occurs in all "reasonable" variational problems and is of great importance for the direct methods.

Consider a function $I[\varphi]$, where the independent function φ ranges over a specific "function space". Consider a sequence of admissible functions φ_n which tend to a limiting function u in this function space. Then I is called lower continuous at the place u if

$$\lim I[\varphi_n] \geq I[u] \quad ,$$

no matter what sequence φ_n tending to u is considered.

Problem

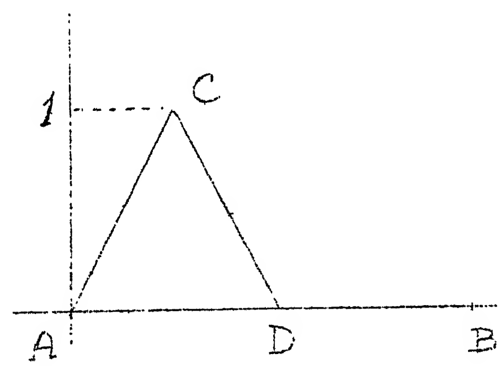
Prove the theorem used above that:

In a sequence of convex curves of bounded length lying in a closed domain there is a subsequence which converges to a closed convex curve.

Compactness In Function Space, Arzela's Theorem and Applications.

In the ordinary theory of maxima or minima, the existence of a greatest or smallest value of a function in a closed domain is assured by Bolzano-Weierstrass convergence theorem: a bounded set of points always contains a convergent subsequence. This fact, together with the continuity of the function, serves to secure the existence of an extreme value.

As seen before, in the calculus of variations the continuity of the function space often has to be replaced by a weaker property, semi-continuity, for, when ϕ_n converges to u, generally $I(\phi_n) \neq I(u)$. Another difficulty in the calculus of variations arises from the fact that the Bolzano-Weierstrass convergence theorem does not hold if the elements of the set are not longer points on a line or in a n-dimensional space, but functions, curves or surfaces.



As an example, we can consider the curves ACDB in a closed interval where $AD = \frac{2}{n}$ and where C remains at the distance 1 from AB. These curves are continuous, but in the sequence there is no subsequence which converges to an admissible continuous curve.

Fortunately, however, there exists a remedy which very often proves sufficient in the direct methods of the calculus of variations. By a suitable restrictive condition imposed on the functions of a set, one can again obtain a theorem analogous to the Bolzano-Weierstrass theorem. This condition is that of equicontinuity.

A set of functions $f_1(p), \ldots f_n(P)$ is <u>equicontinuous</u> in a domain B if, given any $\varepsilon > 0$, there exists a δ, depending on ε

alone, not on the particular $f_n(P)$, such that

$$|f_n(Q)-f_n(P)| < \varepsilon \quad \text{for } |Q-P| < \delta(\varepsilon)$$

uniformly for all P in B and all functions $f_n(\xi)$ This condition implies, of course, that each function separately be continuous.

As an example, let us consider the set of functions

$$\int_0^1 e^{xy} g(y)dy = f(x) ,$$

where $g(y)$ is any piecewise continuous function such that $|g(y)| < 1$ and where x has an upper bound. Then

$$|f(x) - f(\xi)| \leq \int_0^1 |g(y)||e^{xy} - e^{\xi y}|dy ;$$

as

$$|e^{xy} - e^{\xi y}| < \varepsilon$$

for

$$|x-\xi| < \delta(\varepsilon),$$

then

$$|f(x) - f(\xi)| < \varepsilon$$

for $|x-\xi| < \delta(\varepsilon)$, hence $f(x)$ is equicontinuous for all $g(y)$. Of course, e^{xy} can be replaced by any continuous function $K(x,y)$, and the statement of equicontinuity remains true.

As another example, let us consider a sequence $\{f_n(x)\}$ in an interval $(x_0 x_1)$ for which

$$\int_{x_0}^{x_1} [f_n'(x)]^2 dx \leq M ,$$

where M is a fixed constant. $\{f_n(x)\}$ is equicontinuous, since the

expression

$$|f_n(x+h)-f_n(x)|^2 = \left|\int_x^{x+h} f'(x)dx\right|^2$$

by Schwartz' inequality*, does not exceed

$$h\int_x^{x+h} [f'(x)]^2 dx \leq Mh,$$

and we have merely to take, for a given ε,

$$h < \delta(\varepsilon) = \frac{\varepsilon^2}{M} .$$

Arzela's convergence theorem then states: <u>Given in a closed domain B a set of functions $\{f(P)\}$ which are uniformly bounded (i.e. $|f(P)| < M$) and equicontinuous, then there exists a subsequence $\{f_n(P)\}$ which converges uniformly.</u> We cover the domain B with a net Σ_1 of lattice points, and let $\frac{1}{2}$ be the common distance between two consecutive points.

The values of f(P) at the first lattice point form a bounded, infinite set of numbers, and hence, according to the Bolzano-Weierstrass Theorem, there exists a sequence of these values which converges at the first lattice point. From this sequence we can choose a subsequence of functions which converges at the second lattice point (the lattice points can be ordered), and so forth for all lattice points (there is a finite number of them). We thus obtain a subsequence of functions converging at all the lattice points:

$$S_1(P): f_{1,1}(P),\ f_{2,1}(P),\ \ldots$$

---

* Schwartz' inequality for integrals states that

$\left[\int_a^b fg\right]^2 \leq \int_a^b f^2 \int_a^b g^2$, and is proved from the relation

$$\int_a^b (f-g)^2 \geq 0 .$$

Let us now take the middle points between the lattice points and consider the new set of lattice points thus obtained, $\mathfrak{N}_2$. We similarly have a subsequence of functions converging at all these lattice points:

$$S_2(P): \; f_{1,2}(P),\; f_{2,2}(P),\; \ldots$$

Continuing this process, we obtain the following sequences:

$$S_1(P): \; f_{1,1}(P),\; f_{2,1}(P),\; \ldots,\; f_{n,1}(P),\; \ldots$$

$$S_2(P): \; f_{1,2}(P),\; f_{2,2}(P),\; \ldots,\; f_{n,2}(P),\; \ldots$$

$$S_i(P): \; f_{1,i}(P),\; f_{2,i}(P),\; \ldots,\; f_{n,i}(P),\; \ldots$$

Each sequence $S_i(P)$ is a subsequence of all the preceding, and it converges at all the lattice points of the net $\mathfrak{N}_i$.

We now choose the diagonal sequence $\{f_n(P)\}$:

$$f_{1,1}(P),\; f_{2,2}(P),\; \ldots,\; f_{n,n}(P),\; \ldots$$

which, except for a finite number of its terms, is an infinite subsequence of every $S_i(P)$, and hence $\{f_n(P)\}$ converges at all lattice points.

There remains to show that the diagonal subsequence $\{f_n(P)\}$ converges uniformly at any arbitrary point Q of B. ε being given, since the diagonal sequence $\{f_n(P)\}$ is equicontinuous, there exists, uniformly for every k, a $N(\varepsilon)$ such that

$$|f_k(Q) - f_k(L)| < \frac{\varepsilon}{3}$$

for $|Q-L| < 2^{-N(\varepsilon)}$, L being a lattice point of the net $\mathfrak{N}_N$.

Since $\{f_n(P)\}$ converges at each of the lattice points, we can find a $\ell(\varepsilon)$ such that

$$|f_n(L) - f_m(L)| < \frac{\varepsilon}{3}$$

for $m, n > K(\varepsilon)$.

Then, for any point P

$$|f_n(P) - f_m(P)| \leq |f_n(P) - f_n(L)|$$
$$+ |f_m(P) - f_m(L)|$$
$$+ |f_n(L) - f_m(L)|$$

or

$$|f_n(P) - f_m(P)| \leq \varepsilon .$$

Therefore $\{f_n(P)\}$ converges uniformly.

Problem

Prove that, if a set of functions $f_n(P)$ is equicontinuous in a closed domain, and if the $f_n(P)$ are bounded at least at one point, then the functions f_n are bounded everywhere in the domain.

Application to geodesics: Lipschitz's condition: A set functions are said to satisfy Lipschitz's condition if

$$\left| \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right|$$

is uniformly bounded for all functions $f(x)$, provided that x_1 and x_2 are in a closed interval.

According to Hilbert, the preceding concept immediately permits proof of the existence of a shortest connection between two points A and B on a given surface.

Let $[x(t), y(t), z(t)]$ be the parametric representation of a set of curves through two points A and B, t being the arc length or proportional to the arc length. We also assume that the

length of the curves has an upper bound, and we can take $0 \le t \le 1$. Then the functions $x(t)$, $y(t)$, $z(t)$ are equicontinuous.

If we assume that the functions possess piecewise continuous derivatives, $\dot{x}(t)$, $\dot{y}(t)$, $\dot{z}(t)$, we have, t being the arc length or proportional to the arc length,

$$\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} = C .$$

s and σ being two values of t,

$$|x(s) - x(\sigma)| \le \int_{\sigma}^{s} |\dot{x}(t)|\,dt = \int_{\sigma}^{s} \sqrt{[\dot{x}(t)]^2}\,dt ,$$

which implies

$$|x(s) - x(\sigma)| \le \int_{\sigma}^{s} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}\quad dt ,$$

or

$$|x(s) - x(\sigma)| \le C|s - \sigma| ,$$

which means that, under the assumption made for t, the curves are equicontinuous.

But, even if the derivatives do not exist, the conclusion remains the same, as it can easily be seen by representing the length as the lowest upper bound of the sum

$$\sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} ,$$

where the least upper bound is taken with respect to modes of inserting the intermediate points (x_i, y_i, z_i) between A and B on the curve, and for all n.

As the length of the curves has an upper bound, $x(t)$, $y(t)$, and $z(t)$ also have an upper bound, and, as they are equicontinuous there is, by Arzela's theorem, a subsequence $x_m(t)$ such that $x_m(t)$ converges uniformly. Similar subsequences $y_m(t)$ and $z_m(t)$ also converge. Hence there is a limit curve C joining A and B.

By the semi-continuity property of length,

$$L(C) \leq \underline{\lim}\, L(C_m) = d \;,$$

where d is the greatest lower bound of the length of the curves joining A and B on the surface. Hence the < sign is impossible and

$$L(C) = d \;.$$

There remains to show that the minimizing curve C has piecewise continuous first and second derivatives and satisfy Euler's equation, provided that the given surface is sufficiently smooth. But this follows from the fact that C must give the shortest length between any two of its points. If B is sufficiently near A, the arc AB of C must be regular and satisfy Euler's equation on the basis of the classical theory of the calculus of variations (See first part of these notes). Hence this property follows for the entire arc of C between the given end-points.

Problem

Prove by the method above that there exists on a torus a shortest geodesic topologically equivalent to any prescribed closed circuit on the torus.

Direct variational methods in the theory of integral equations.

Following Holmgren, we shall indicate briefly how direct variational methods can be used for the treatment of Fredholm and Hilbert integral equations. In the general theory of integral equations, which was inaugurated by Fredholm. Hilbert emphasized the importance of the eigenvalue theory.

Let $K(s,t)$ be a continuous and symmetric function called the "kernel", with $0 \leq s \leq 1$ and $0 \leq t \leq 1$. The eigenvalue problem is to find a function $u(t)$ such that

$$\int_0^1 K(s,t)u(t) = \mu\, u(s)$$

Such functions u are called eigen functions and the $\lambda = 1/\mu$ eigenvalues.

We shall now prove the existence of one eigen-value. We consider the integral

$$K(\phi,\phi) = \iint K(s,t)\phi(s)\phi(t)dsdt \quad ,$$

where ϕ is piecewise continuous in the interval (0,1). We assume that $|K(s,t)|$ has an upper bound. (Integrals without indication of limits shall be taken over the domain defined by $0 \leq s \leq 1$ and $0 \leq t \leq 1$.) ..

We assume that the kernel $K(s,t)$ is such that $K(\phi,\phi)$ is sometimes positive. in other words, that $K(s,t)$ is not a so-called "negative definite" kernel. We write

$$\int [\varphi(s)]^2 ds = (\varphi,\varphi) \ .$$

Then

$$\frac{K(\phi.\phi)}{(\phi,\phi)}$$

has a least upper bound μ, and there exists a maximizing sequence $\phi_1, \ldots, \phi_n, \ldots$, for which

$$\frac{K(\phi,\phi)}{(\phi,\phi)} \rightarrow \mu \ .$$

The function ϕ could be normed, that is $(\phi,\phi) = 1$, but this is not essential.

We now consider the new sequence $\phi_n + \varepsilon\zeta_n$, and we have

$$K(\phi_n + \varepsilon\zeta_n,\ \phi_n + \varepsilon\zeta_n) \leq \ (\phi_n + \varepsilon\zeta_n,\ \phi_n + \varepsilon\zeta_n)$$

or

$$K(\phi_n,\phi_n) + 2\varepsilon K(\phi_n,\zeta_n) + \varepsilon^2 K(\zeta_n,\zeta_n) \leq$$

$$\mu[(\phi_n,\phi_n) + 2\varepsilon(\phi_n,\zeta_n) + \varepsilon^2(\zeta_n,\zeta_n)] \ ,$$

or

$$K(\phi_n, \phi_n) - \mu(\phi_n, \phi_n) + 2\varepsilon[K(\phi_n, \zeta_n) - \mu(\phi_n, \zeta_n)]$$

$$+ \varepsilon^2[K(\zeta_n, \zeta_n) - \mu(\zeta_n, \zeta_n)] \leq 0 .$$

The quadratic form in will always be negative if

$$[K(\phi_n, \zeta_n) - \mu(\phi_n, \zeta_n)]^2$$

$$-[K(\phi_n, \phi_n) - \mu(\phi_n, \phi_n)]\,[K(\zeta_n, \zeta_n) - \mu(\zeta_n, \zeta_n)] \leq 0 .$$

We assume that (ζ_n, ζ_n) remains bounded, hence, as $K(\phi_n, \phi_n) - \mu(\phi_n, \phi_n) \to 0$,

$$K(\phi_n, \zeta_n) - \mu(\phi_n, \zeta_n) \to 0,$$

or

$$\int \zeta_n(s)[\int K(s,t)\phi_n(t)dt - \mu\phi_n(s)]ds \to 0.$$

If we write

$$\int K(s,t)\phi_n(t)dt = \eta_n(s),$$

the functions $\eta_n(s)$ are equicontinuous and equibounded, hence, by Arzela convergence theorem, there exists a subsequence, which we shall also call $\eta_n(s)$, of these functions converging uniformly to a continuous function u(s).

u(s) is not identically 0. If it were,

$$\int K(s,t)\phi_n(t)dt$$

would tend to 0. $\phi_n(t)$ be assumed to be normed, then the

upper bound μ would not be positive, which is contrary to the hypothesis. Hence u(s) is not identically 0.

With the new notation,

$$\int \zeta_n(s)\left[\int K(s,t)\phi_n(t)dt - \mu\phi_n(s)\right]ds$$

becomes

$$\int \zeta_n(s)\left[\eta_n(s) - \mu\phi_n(s)\right]ds \quad ,$$

and it tends to 0. We consider a special variation for ζ:

$$\zeta_n(s) = K(s,r) \ ,$$

where r is an arbitrary parameter such that

$$0 \leq r \leq 1 \ .$$

Then

$$\int K(s,r)\eta_n(s)ds - \mu\eta_n(r)$$

tends to 0. But $\eta_n(r)$ converges uniformly to u(r), hence

$$\int K(s,r)u(s)ds = \mu u(r) \ .$$

If we multiply both sides by u(r) and integrate, we verify immediately that u(r) is a solution.

By the same method we can solve the inhomogeneous Fredholm equation

$$\int K(s,t)u(t)dt - \nu u(s) - g(s) = 0,$$

where ν is a constant greater than μ, and g(s) a given function.

Solving this problem is equivalent to solving the following variational problem: Find the function u for which the integral

$$I(\phi) = \iint K(s,t)\phi(s)\phi(t)dsdt - \nu\int[\phi(s)]^2ds$$
$$- 2\int g(s)\phi(s)ds$$

is a maximum.

The proof is left to the student.

Dirichlet's Principle

irichlet's integral. Let G be a domain of the xy-plane, the oundary of which, , is a Jordan curve, i.e., a continuous curve ithout double points. $\varphi(x,y)$ is continuous in $G+\gamma$, φ_x and φ_y eing piecewise continuous in G. $\varphi = \overline{g}$ on γ, $\overline{g}$ being continuous n γ. Functions φ satisfying these conditions form a class of ad- ssible functions. Dirichlet's integral for φ is defined as:

$$D(\varphi) = \iint_G (\varphi_x^2 + \varphi_y^2)\,dx\,dy = \iint_G (\varphi_r^2 + \frac{1}{r^2}\varphi_\theta^2)\,r\,dr\,d\theta \text{ .}$$

φ) has a greatest lower bound d, and

$$D(\varphi) \geq d$$

Minimizing sequences. For a variational integral $I(\varphi)$ a sequence f admissible functions $\varphi_1, \varphi_2, \varphi_3, \ldots$ such that the values of (φ_1), $I(\varphi_2)$, $I(\varphi_3)$, ... tend to the greatest lower bound d of (φ) is called a minimizing sequence. Wherever the set of possible alues of $I(\varphi)$ is bounded from below, the existence of a minimizing equence is insured, even though we need not specify a definite consti onstruction of the minimizing sequence. (Such a construction will o the main point in the task (c) of computing numerical values.) or example, Dirichlet's integral having a greatest lower bound d, here is a minimizing sequence of functions $\varphi_n(x,y)$ for which $(\varphi_n) \to d$ as $n \to \infty$. However, a minimizing sequence need not onverge; even if it converges the limit function need not be ad- issible.

xplicit expression of Dirichlet's integral for a circle. Hadamard's bjection. The difficulties just mentioned are clearly shown by a act first discovered by Hadamard: not only is the solvability of irichlet's variational problem not obvious, but Dirichlet's inimum problem actually is unsolvable in some cases in which the oundary value problem for the differential equation $\Delta u = 0$ can

be solved. Thus the idea of reducing the latter to the former seemed even more discredited.

Let K be the unit circle, and introduce polar coordinates r, θ. On the circumference k of K continuous boundary values $\bar{g} = \bar{g}(\theta)$ are given. Consider the (not necessarily convergent) Fourier series of $\bar{g}$:

$$\frac{a_o}{2} + \sum_{\nu=1}^{\infty}(a_\nu \cos\nu\theta + b_\nu \sin\nu\theta) .$$

Then, for r < 1, the solution of $\Delta u = 0$ satisfying the boundary condition $\bar{u} = \bar{g}$ on k is given by the convergent series:

$$u(r,\theta) = \frac{a_o}{2} + \sum_{\nu=1}^{\infty} r^\nu (a_\nu \cos\nu\theta + b_\nu \sin\nu\theta).$$

If we represent by $D_\rho(u)$ Dirichlet's integral for the circle of radius $\rho < 1$ about the origin,

$$D_\rho(u) = \pi\sum_{\nu=1}^{\infty}\nu(a_\nu^2 + b_\nu^2)\rho^{2\nu} .$$

This implies for every N

$$\pi\sum_{\nu=1}^{N} \nu(a_\nu^2 + b_\nu^2)\rho^{2\nu} \le D(u) \le \pi\sum_{\nu=1}^{\infty}\nu(a_\nu^2 + b_\nu^2) ,$$

where the right-hand side may be a divergent series. By letting ρ tend to 1, we infer immediately: Dirichlet's integral for the harmonic function

$$u(r,\theta) = \frac{a_o}{2} + \sum_{\nu=1}^{\infty} r^\nu(a_\nu \cos\nu\theta + b_\nu \sin\nu\theta)$$

over the unit circle is given by the series

$$D(u) = \pi\sum_{\nu=1}^{\infty}\nu(a_\nu^2 + b_\nu^2)$$

and exists if, and only if, this series converges.

Now, as pointed out by Hadamard, there exist continuous functions $\overline{g}(\theta)$ for which this series diverges; e.g. if $\overline{g}(\theta)$ is given by the uniformly convergent Fourier expansion

$$\overline{g} = \sum_{\mu=1}^{\infty} \frac{\sin \mu! \theta}{\mu^2} \quad .$$

hen

$$D(u) = \pi \sum_{\mu=1}^{\infty} \frac{\mu!}{\mu^4} \quad ,$$

hich does not converge. With boundary values such as this $\overline{g}$, the oundary value problem of $\Delta u = 0$ can therefore certainly not be educed to a variational problem for Dirichlet's integral, and irichlet's principle is invalid. No full equivalence between the ariational problem and the boundary value problem exists.

<u>ne Correct Formulation of Dirichlet's Principle.</u> The last ifficulty can be avoided by restricting the prescribed boundary alue in such a manner as not to exclude from the outset the olvability of the variational problem. While for the boundary alue problem such conditions are not necessary, they are essential o make the variational problem meaningful. Accordingly, it will e assumed that the prescribed boundary values $\overline{g}$ are the values on of a function g in $G + \gamma$ for which $D[g]$ is finite. In other ords, we explicitly assume that there exists at least one admissible function with a finite Dirichlet integral. Thus one is d to the following formulation:

<u>richlet's Principle.</u> Given a domain G whose boundary is a Jordan rve. Let g be a function continuous in $G + \gamma$, piecewise smooth in and with a finite Dirichlet integral $D(g)$. Let φ be the class all functions continuous in $G + \gamma$, piecewise smooth in G, and th the same boundary values as g. Then the problem of finding a nction for which $D(\varphi)$ = minimum = d, has a unique solution $\varphi = u$. is function u is the solution of the boundary value problem $\Delta u = 0$ with the values $\overline{g}$ on γ.

<u>Lower semi-continuity of Dirichlet's integral for harmonic functions.</u>

If a sequence of harmonic functions u_n converges to a harmonic function u uniformly in every closed subdomain of G, then

$$D_G(u) \leq \underline{\lim}\, D_G(u_n) .$$

<u>Proof:</u> For any closed subdomain G' of G the assumed uniform convergence of the u_n implies by Harnack's theorem* the uniform convergence of the derivatives of u_n to those of u. Hence

$$D_{G'}(u) = \lim_{n \to \infty} D_{G'}(u_n) \leq \underline{\lim}\, D_G(u_n)$$

and, by letting G' tend to G,

$$D_G(u) \leq \underline{\lim}\, D_G(u_n) .$$

<u>roof of Dirichlet's Principle for the Circle.</u> Let the domain G o the unit circle and consider the Fourier series of the given oundary function $\bar{g} = g(1, \theta)$

$$\frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu\theta + b_\nu \cos \nu\theta) .$$

his series need not converge, but for r < 1 the series

$$u = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} r^2 (a_\nu \cos \nu\theta + b_\nu \cos \nu\theta)$$

oes converge, and u is harmonic.

Let v be any other admissible function for which D(v) is inite, and $\zeta = u - v$. On the boundary $\zeta = 0$.

Harnack's theorem states: If a sequence of harmonic functions onverges uniformly in a domain, then their derivatives converge niformly in every closed subdomain and the limit function is again armonic.

a proof, see, e.g., <u>Foundations of Potential Theory</u> by Kellogg, p. 248.

$$D(v) = D(u - \zeta) = D(u) + D(\zeta) - 2D(u,\zeta),$$

where $D(u,\zeta)$ represents the so-called bilinear form

$$D(u,\zeta) = \iint_G (u_x \zeta_x + u_y \zeta_y)\,dxdy \ .$$

By Green's formula

$$D(u,\zeta) = -\iint_G \zeta \Delta u \; dxdy + \int_\gamma \zeta \frac{\partial u}{\partial r}\, ds \ .$$

The first term of the right side vanishes because $\Delta u = 0$, and the last term vanishes too because $\zeta = 0$ on γ. Hence

$$D(u,\zeta) = 0$$

and, as

$$D(\zeta) > 0 \ ,$$

$$D(v) \geq D(u) \ , \quad \text{which}$$

means that Dirichlet's integral is minimum for u.

This reasoning, however, has a gap, for Green's formula is not applicable to u in the whole domain. In order to fill this gap, consider the "harmonic polynomials"

$$u_n = \frac{a_0}{2} + \sum_{\nu=1}^{n} r^2 (a_\nu \cos\nu\theta + b_\nu \sin\nu\theta)$$

and write $\zeta_n = u_n - v$. Then

$$D(v) = D(u_n) + D(\zeta_n) - 2D(u_n,\zeta_n) \ .$$

By Green's formula, applicable to the polynomials u_n,

$$D[u_n,\zeta_n] = -\iint_G \zeta_n \Delta u_n \, dxdy + \int_{\substack{0 \\ r=1}}^{2\pi} \zeta_n \frac{\partial u_n}{\partial r}\, d\theta$$

The first term of the right side vanishes because $\Delta u_n = 0$. It can be seen that the second term vanishes too by substituting for u_n its explicit expression, and by observing that the first 2n + 1

$D(v - u) = 0$, i.e., $v - u = \text{const.}$ But u and v having the same boundary values

$$v - u = 0 .$$

<u>"Distance" in function space.</u> <u>Triangle inequalities.</u> Let φ and ψ be two admissible functions, and write

$$D(\varphi,\psi) = \iint_G (\varphi_x\psi_x + \varphi_y\psi_y)\, dxdy \quad .$$

With this notation

$$D(\varphi) = D(\varphi,\varphi)$$

and

$$D(\varphi+\psi) = D(\varphi) + D(\psi) + 2D(\varphi,\psi) .$$

As $D(\varphi)$ has a greatest lower bound, there exists a minimizing sequence $\varphi_1, \varphi_2, \dots$ Consider the new sequence $\varphi_n + \varepsilon\zeta_n$. These functions will be admissible if ζ_n and its derivatives satisfy certain conditions of continuity and if ζ_n vanishes at the boundary.

$$D(\varphi_n + \varepsilon\zeta_n) = D(\varphi_n) + 2\varepsilon D(\varphi_n,\zeta_n) + \varepsilon^2 D(\zeta_n) \geq d \quad .$$

We assume that $D(\zeta_n)$ has an upper bound R. If we write $D(\varphi_n) = d_n$ and $D(\varphi_n,\zeta_n) = V_n$, the above inequality can be written

$$D(\varphi_n + \varepsilon\zeta_n) = d_n + 2\varepsilon V_n + \varepsilon^2 D(\zeta_n) \geq d$$

or, a fortiori,

$$d_n + 2\varepsilon V_n + \varepsilon^2 R(\zeta_n) \geq d$$

or

$$\varepsilon^2 R(\zeta_n) + 2\varepsilon V_n + d_n - d \geq 0 .$$

This inequality is satisfied for any ε, if the quadratic form in is positive definite, that is if

$$V_n^2 - R(d_n - d) \leq 0 ,$$

which implies that

$$V_n = D(\varphi_n, \zeta_n) \text{ tends uniformly to } 0 \text{ as } n \to \infty .$$

We now take a special function $\zeta_n = \phi_n - \phi_m$, where m is fixed. Then

$$|D(\phi_n, \phi_n - \phi_m)| \leq \sqrt{R(d_n - d)}$$

and also

$$|D(\phi_m, \phi_m - \phi_n)| \leq \sqrt{R(d_m - d)} .$$

By subtracting the left-hand sides,

$$D(\phi_m - \phi_n) \leq \sqrt{R(d_m - d)} + \sqrt{R(d_n - d)} .$$

Hence

$$D(\phi_m - \phi_n) \to 0 \text{ as } m, n \to \infty .$$

$D(\phi - \psi)$ can be considered as the "distance" between two functions ϕ and ψ in the function space under consideration. For these "distances" the following so-called triangle inequalities hold:

$$\sqrt{D(\phi)} + \sqrt{D(\psi)} \geq \sqrt{D(\phi + \psi)}$$

$$\sqrt{D(\phi)} - \sqrt{D(\psi)} \leq \sqrt{D(\phi - \psi)}$$

The fact that $D(\phi - \phi_n) \to 0$ as $n \to \infty$ does not imply that ϕ_n converges to ϕ. If we take $\phi = 0$, the following example will show that, although $D(\phi_n) \to 0$ as $n \to \infty$, ϕ_n need not tend to 0.

Consider the minimum problem for $D(\varphi)$ in a circle of radius 1, when the admissible functions are to vanish on the boundary. This minimum problem is solved by $\varphi \equiv 0$ and by no other function; $d = 0$ is the minimum value, not merely the greatest lower bound.

Now we define in polar coordinates r, θ a sequence of admissible functions φ_n by

$$\varphi_n = \begin{cases} 1 & , \text{ for } r \leq \rho_n^2 \\ \dfrac{\log r}{\log \rho} - 1 & , \text{ for } \rho_n^2 \leq r \leq \rho_n \\ 0 & , \text{ for } \rho_n \leq r \leq 1 \end{cases} .$$

Then

$$D(\varphi_n) = \int_0^{2\pi} \int_{\rho_n^2}^{\rho_n} \left(\frac{\partial \varphi_n}{\partial r}\right)^2 r \, dr \, d\theta = 2\pi \int_{\rho_n^2}^{\rho_n} \left(\frac{1}{r \log \rho_n}\right)^2 r \, dr = - \frac{2\pi}{\log \rho_n} .$$

We now take $\rho_n = \frac{1}{n}$. Then

$$D(\varphi_n) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty ,$$

but φ_n does not tend to 0 at the center of the circle, where its value is 1.

Thus a minimizing sequence cannot in general be expected to yield the solution of the problem by a mere passage to the limit. The essential point in the "direct variational methods" is to introduce an appropriate sequence that will guarantee convergence.

<u>Construction of an harmonic function u by a "smoothing process"</u>: Consider a minimizing sequence φ_n of admissible functions in G. As seen by the example above, the functions φ_n need not converge.

We consider a circle K in G, and we replace a function φ_n by a function ω_n such that $\omega_n = \varphi_n$ outside of K and ω_n is harmonic in K. By this operation we are "smoothing out" φ_n. Dirichlet's principle having been proved for the circle,

$$D_k(\omega_n) \leq D_k(\varphi_n) .$$

and therefore

$$D(\omega_n) \leq D(\varphi_n) .$$

The functions ω_n are admissible functions and form a new minimizing sequence. Hence

$$D(\omega_n - \omega_m) \to 0 \quad \text{as } n \text{ and } m \to \infty .$$

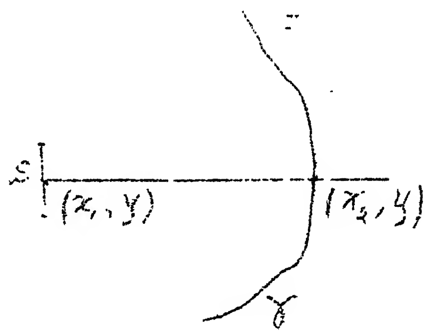
Let us write

$$\omega_n - \omega_m = \sigma_{nm} .$$

We can find an N such that in a concentric circle K' in K

$$D_{K'}(\sigma_{nm}) < \varepsilon^2 \text{ for } n,m > N .$$

Through a point (x_1,y) in K' we consider the parallel to the x-axis, and let (x_2,y) be its intersection with the boundary γ of G. As $\sigma_{nm}(x_2,y) = 0$, we can write



$$\sigma_{nm}(x_1,y) = \sigma_{nm}(x_1,y) - \sigma_{nm}(x_2,y) = \int_{x_2}^{x_1} \sigma_{nm_x}(x,y)dx ,$$

which implies by Schwarz's inequality

$$\sigma^2_{nm}(x_1,y) \leq |x_1 - x_2| \int_{x_2}^{x_1} \sigma^2_{nm_x}(x,y)dx ,$$

or, as G is bounded,

$$\sigma^2_{nm}(x_1,y) \leq L \int_{x_2}^{x_1} \sigma^2_{nm_x}(x,y)dx ,$$

where L is a given length. By integrating between two values of y, y_1 and y_2, in K',

$$\int_{y_2}^{y_1} \sigma_{nm}^2(x_1, y)dy < L\int_{y_2}^{y_1}\int_{x_2}^{x_1}\sigma_{nm_x}^2(x,y)dxdy$$

or, a fortiori,

$$\int_{y_2}^{y_1}\sigma_{mn}^2(x,y)dy < L\ D_{K'}(\sigma_{mn}) .$$

The integral of the left-hand side is taken along a segment of straight line s in K' parallel to the y-axis. Hence at some point of s

$$\sigma_{nm}^2 < \frac{L\varepsilon^2}{|y_2 - y_1|} .$$

Hence $\sigma_{nm} = \omega_n - \omega_m$ tends uniformly to 0, and ω_n tends uniformly to a function u. u is an harmonic function because of the following theorem (which will be proved later on in these notes):

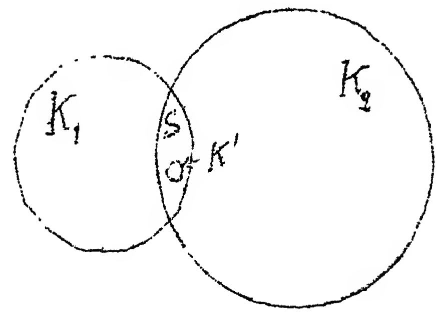
If in a circle a sequence of harmonic functions converges to a limit, the limit is an harmonic function.

If we let now K' tend to K, we have defined a harmonic function u in K.

The smoothing process can be applied to any circle in G; it leads in every such circle to the definition of a certain harmonic function u. We assert that this construction defines a uniquely determined function in the whole domain G. For the proof we need only show that the functions u_1 and u_2 resulting from the smoothing in two overlapping circles K_1 and K_2 are identical in the common part S of these two circles.

Let ω_n^1 and ω_n^2 be the minimizing sequences originating from the sequence φ_n by smoothing in the circle K_1 and K_2 respectively. Then the mixed sequence $\omega_1^1, \omega_1^2, \omega_2^1, \omega_2^2, \ldots$ is also a minimizing sequence, and therefore $D_s(\omega_n^1 - \omega_n^2) \to 0$ as $n \to \infty$. If K' is

a circle in S, a fortiori, $D_{K'}(\omega_n^1 - \omega_n^2)$ $\quad 0$. ω_n^1 and ω_n^2 are harmonic in K' and converge to u_1 and u_2 respectively. From the preceding argument it follows that the mixed sequence also converges to a harmonic function u' in K'. Therefore u_1 and u_2 are identical with u' in K'.



<u>Proof that D(u) = d.</u> Let G' be a closed subdomain of G. It can be covered by a finite number of circles K_i, and by taking the circles small enough, we can be sure that they all lie entirely in G. Then

$$G' \subseteq \sum_i K_i \subset B ,$$

and hence

$$D_{G'}(\varphi_n - u) \leq \sum_i D_{K_i}(\varphi_n - u) .$$

By the triangle inequality

$$\sqrt{D_{G'}(u)} \leq \sqrt{D_{G'}(\varphi_n - u)} + \sqrt{D_{G'}(\varphi_n)} \leq \sqrt{D_{G'}(\varphi_n - u)} + \sqrt{D(}$$

The left side of this inequality is independent of n, therefore

$$\sqrt{D_{G'}(u)} \leq \text{g. l. b.} \left(\sqrt{D_{G'}(\varphi_n - u)} + \sqrt{D(\varphi_n)}\right) ,$$

the g.l.b. being taken over all values of n. But

$$D_{G'}(\varphi_n - u) \to 0 \text{ and } D(\varphi_n) \to d \text{ as } n \to \infty ,$$

therefore

$$\sqrt{D_{G'}(u)} \leq \sqrt{d}, \text{ i.e. } D_{G'}(u) \leq d .$$

Consequently, since

$$D(u) = \lim_{G' \to G} D_{G'}(u) ,$$

$$D(u) \leq d ,$$

However, u being an admissible function,

$$D(u) \geq d ,$$

hence

$$D(u) = d .$$

Proof that the function u attains the prescribed boundary values. The function u, so far defined in G only, can be extended continuously to γ and assumes the prescribed boundary values $\bar{g}$ on γ. More precisely: there is a quantity $\varepsilon(\delta)$ tending to 0 with δ such that for all points P in G we have $|u(P) - g(R)| < \varepsilon$ if the distance PR from P to a point R on γ is less than δ. This statement expresses the uniform convergence of u(P) to the prescribed boundary values.

For the proof we assume that R is one of the points on γ nearest to P, at the distance $PR = 5h < \delta$.

We now consider the circle of radius 10h about R. This circle defines in G a certain subregion L.

We shall show first that h can be taken so small that

$$D_L(\varphi_n) < \sigma_1^2(h) \qquad \text{for } \underline{\text{all}}\ n .$$

As $D(\varphi_n - \varphi_m) \to 0$, we can choose N so large that

$$D_L(\varphi_n - \varphi_m) \leq D(\varphi_n - \varphi_m) < \frac{\sigma_1^2}{4}$$

for $n, m \geq N$; and we can choose h so small that

$$D_L(\varphi_\nu) < \frac{\sigma_1^2}{4} \text{ for } \nu \leq N .$$

By the triangle inequality

$$\sqrt{D_L(\varphi_n)} \leq \sqrt{D_L(\varphi_n - \varphi_N)} + \sqrt{D_L(\varphi_N)} \leq \frac{\sigma_1^2}{4} + \frac{\sigma_1^2}{4} < \sigma_1^2 .$$

We now make use of the following

<u>Lemma:</u> There is a circle $r = \hat{r}$ about P, with $3h \leq \hat{r} \leq 4h$, such that for two points M_1 and M_2 on each connected arc of this circle the inequality

$$|\varphi(M_2) - \varphi(M_1)|^2 \leq \frac{2\pi\hat{r}\sigma_1^2}{h} \leq 8\pi\sigma_1^2 ,$$

where σ_1^2 is an upper bound of $D_L(\varphi_n)$, holds.

In polar coordinates this last fact can be written

$$\iint_L (\varphi_r^2 + \frac{1}{r^2}\varphi_\theta^2) r \, d\theta dr \leq \sigma_1^2 ,$$

or, by writing $r\theta = s$, a fortiori

$$\int_{3h}^{4h} dr \int \varphi_s^2 ds \leq \sigma_1^2 .$$

By the mean value theorem for integrals there will be a value $r = \check{r}$ with $3h \leq \check{r} \leq 4h$, for which

$$\int_0^{2\pi r} \varphi_s^2 ds \leq \frac{\sigma_1^2}{h} .$$

On this circle $r = \hat{r}$ we have for the oscillation of φ between two points M_1 and M_2

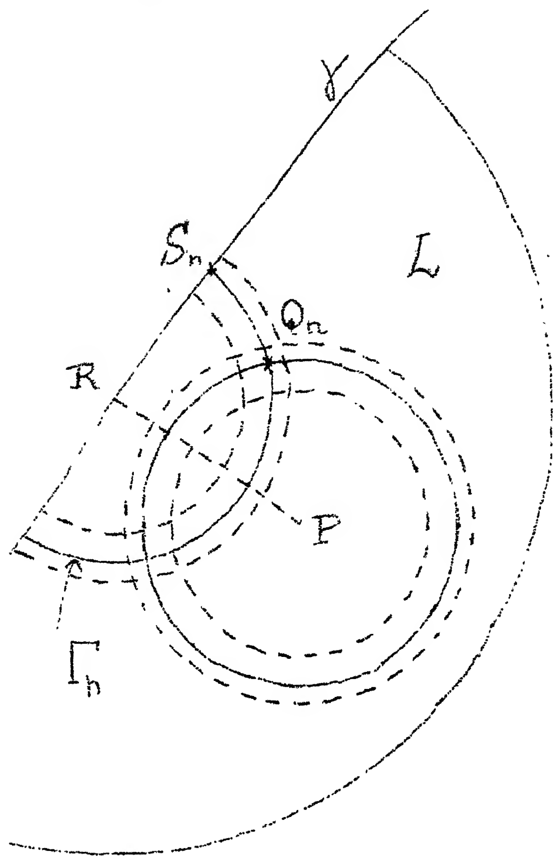
$$|\varphi(M_2) - \varphi(M_1)|^2 \leq 2\pi\check{r}\,\frac{\sigma_1^2}{h} \leq 8\pi\sigma_1^2 ,$$

which proves the lemma. This can be written

$$\varphi_n(M_2) - \varphi_n(M_1) \leq 2\sqrt{2\pi}\,\sigma_1(h)$$

where $\sigma_1(h)$ is independent of n.

Similarly, there exists a circle Γ_n about R with radius 3h on whose circumference the oscillation of φ_n is again less than $2\sqrt{2\pi}\,\sigma_1(h)$. This circle will intersect the circle $\mathcal{E}$ about P at a point Q_n and if h is sufficiently small, it will also intersect γ at a point S_n so that and are connected by an arc of this circle in G. Therefore, since $\overline{\varphi}_n = \overline{g}$,



$$|\varphi_n(Q_n) - g(S_n)| \leq 2\sqrt{2\pi}\,\sigma_1(h).$$

Let u_n be the harmonic function in the circle C_n about P obtained by smoothing φ_n; then the value $u_n(P)$ coincides with some value of φ_n on the circle C_n and hence cannot differ from $\varphi_n(Q_n)$ by more than $2\sqrt{2\pi}\,\sigma_1(h)$:

$$|u_n(P) - \varphi_n(Q_n)| \leq 2\sqrt{2\pi}\,\sigma_1(h) .$$

Finally, since g is continuous on γ, h can be taken such that

$$|g(S_n) - g(R)| \leq \sigma_2(h) ,$$

where $\sigma_2(h) \to 0$ as $h \to 0$.

By combining the three inequalities

$$|u_n(P) - g(R)| \leq 4\sqrt{2\pi}\,\sigma_1(h) + \sigma_2(h) ,$$

which proves that u attains the boundary values.

Thus u is recognized as the solution of the boundary value problem. Since it has been proved that $D(u) = d$ and since u is admissible in the variational problem, it solves the variational problem. The proof of the uniqueness of this solution of the variational problem is now exactly the same as in the case of the circle. Hence Dirichlet's principle, as we stated it above, is established.

<u>Mean value property of harmonic functions</u>. We shall be able to give an alternative proof of Dirichlet's principle through the characterization of harmonic functions by their mean value property.

Let $u(x,y)$ be a continuous function, with continuous first and second derivatives, satisfying the differential equation

$$u_{xx} + u_{yy} = 0 ,$$

then

$$u(P) = \frac{1}{2\pi}\int_C u(\overline{P})\,d\theta ,$$

where the integral is taken along a circle C about P.

By Gauss's theorem

$$\int_C (u_{xx} + u_{yy})\,dx\,dy = \int_C \left(u_x \frac{\partial x}{\partial r} + u_y \frac{\partial y}{\partial r}\right) ds = 0$$

or

$$\int_C \frac{\partial u(\overline{P})}{\partial r}\, r\,d\theta = 0 ;$$

or, because of the continuity of the first derivatives of u,

$$\frac{d}{dr}\int_C u(\overline{P})\,d\theta = 0 ,$$

which means that

$$\frac{1}{2\pi}\int u(P)d\theta = \text{const. with respect to } r.$$

If we let r tend to 0, we see that the value of the constant is u(P), hence

$$u(P) = \frac{1}{2\pi}\int_C u(\overline{P})d\theta \, .$$

We shall now prove the converse, namely, that if a function, which is assumed to be continuous, satisfies the mean value equality for any radius, then the function is harmonic.

$$u(P) = \frac{1}{2\pi}\int_0^{2\pi} u(\overline{P})d\theta$$

for every r.

We multiply both sides by r and integrate with respect to r. Hence

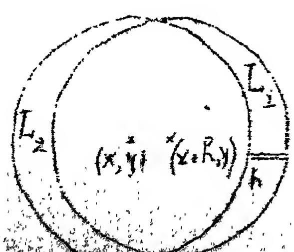
$$u(P) = \frac{1}{\quad^{2}}\iint u(\overline{P})dxdy \ ,$$

which can be considered as the mean value theorem for the area (the double integral is taken over the circle of radius r about P)

In order to prove that u is differentiable, we consider the expression:

$$\pi r^2\,\frac{u(x+h,\ y) - u(x,y)}{h} = \frac{1}{h}\iint_{L_1} u(\overline{P})dxdy - \frac{1}{h}\iint_{I} u(\overline{P})dxdy \ ,$$

where the first integral is taken over the lune L_1 and the second over the lune L_2 between the circles of radius r about the points (x,y) and (x+h,y).



$$\frac{1}{h}\iint_{L_1} u(\overline{P})dxdy = \frac{1}{h}\int_{-r}^{r} dy\int_{0}^{h} u(\overline{P})dx = \int_{-r}^{r} u(\check{P})dy \ ,$$

where $\check{P}$ is some point in L_1 on the segment parallel to the x-axis. Because of the continuity of the function u,

$$\int_{-r}^{r} u(\check{P})dy \rightarrow \int_{-r}^{r} u(\overline{P})dy \quad \text{as } h \rightarrow 0 \ ,$$

and similarly for the other lune. Hence, as $h \rightarrow 0$,

$$\frac{1}{h}\iint_{L_1} u(\overline{P})dxdy - \frac{1}{h}\iint_{L_2} u(\overline{P})dxdy \rightarrow \int_C u(\overline{P})dy,$$

where the integral in the right-hand side is taken along the circle C of radius r about the point (x,y). Hence

$$\pi r^2 \ \frac{u(x+h,y) \ -x(x,y)}{h} \rightarrow \int_C u(\overline{P})dy \quad \text{as } h \rightarrow 0 \ ,$$

and u possesses a derivative u_x given by the integral

$$\frac{1}{\pi r^2}\int_C u(\overline{P})dy \ .$$

By Gauss's theorem,

$$u_x = \frac{1}{\pi r^2}\int_C u(\overline{P})dy = \frac{1}{\pi r^2}\iint_C u_x dxdy \ ,$$

which means that u_x possesses the same mean value property as u. Hence the function u has continuous derivatives of every order.

We shall now prove that u is harmonic. Since u has continuous derivatives, we can write

$$u(x+h\cos\theta, \ y + h\sin\theta) = u(x,y) + h\cos\theta\, u_x + h\sin\theta\, u_y$$
$$+ h^2(u_{xx}\cos^2\theta + 2u_{xy}\sin\theta\cos\theta + u_{yy}\sin^2\theta) + h^3R \ ,$$

where R is bounded.

Integrating both sides with respect to θ along the circle C of radius h about P, we obtain

$$\int_C u(\overline{P})d\theta = 2\pi\ u(x,y) + \pi h^2(u_{xx} + u_{yy}) + h^3 M ,$$

where M is bounded. As the mean value property has been assumed for the function u,

$$\pi h^2(u_{xx} + u_{yy}) + h^3 M = 0 ,$$

or

$$(u_{xx} + u_{yy}) + hR = 0,$$

and, by letting $h \to 0$,

$$u_{xx} + u_{yy} = 0 ,$$

which proves that u is harmonic.

As an application, we can prove that if a sequence of harmonic functions u_n converges uniformly to u, then the limit function u is harmonic. (We have made use of that theorem previously.)

$$u = \lim_{n\to\infty} u_n = \lim_{n\to\infty} \frac{1}{2\pi}\int_0^{2\pi} u_n(\overline{P})d\theta ;$$

but, because of the uniform convergence, we can interchange the order of the operations of integration and passing to the limit, and we obtain

$$u = \frac{1}{2\pi}\int_0^{2\pi} \lim_{n\to\infty} u_n(\overline{P})d\theta = \frac{1}{2\pi}\int_0^{2\pi} u(\overline{P})d\theta ,$$

hence u is harmonic.

Alternative proof of Dirichlet's principle: We make the same assumption as at the beginning of the first proof, and we consider a minimizing sequence of admissible functions $\varphi_1, \varphi_2, \dots, \varphi_n, \dots$

At a point P we consider the circle C of radius h about P. This circle will lie entirely in G for all points P whose distance

from the boundary is greater than h. At P we consider the function

$$\omega_n(P) = \frac{1}{\pi h^2} \iint_{a} \varphi_n(\overline{P})dxdy .$$

The functions φ_n being continuous, the functions ω_n are differentiable.

We shall now prove that, for a fixed h, $\omega_n(P)$ tends uniformly to $u(P)$.

For the admissible functions

$$D(\varphi_n - \varphi_m) \to 0 \quad \text{as } n,m \to \infty$$

and

$$D(\varphi_n, \zeta_n) \to 0 \quad \text{uniformly as } n \to \infty$$

provided that $D(\zeta_n)$ remains bounded, the functions ζ_n have piecewise continuous first derivatives, and $\overline{\zeta}_n = 0$.

Let us write

$$H(\zeta) = \iint_G \zeta^2 dxdy.$$

We will prove that

$H(\zeta) < CD(\zeta)$, where C is a positive constant.

We remark that we need to prove the proposition only for a square. G being bounded, we can surround it by a square and consider a function ζ^* such that

$$\zeta^* = \zeta \text{ in } G$$

and

$$\zeta^* = 0 \quad \text{outside of } G \text{ but}$$

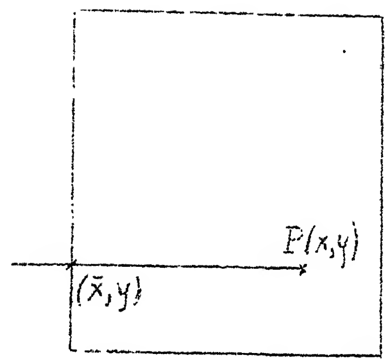
inside the square surrounding G.

Let P(x,y) be a point in a square, we take x and y-axes parallel to the sides of the square and let $\overline{x}$ the point where the parallel to the x-axis through P intersects the side of the square.

Then, as $\zeta(\overline{x},y) = 0$,

$$\zeta(x,y) = \int_{\overline{x}}^{x} \zeta_x dx.$$

(We remark that the parallel to the x-axis through P must cut the boundary of G at a finite number of points at most, in order that the discontinuity of the derivative ζ_x occurs at a finite number of points at most. This is the only of G.)



By Schwarz's inequality,

$$\zeta(P) = \int_{\bar{x}}^{x} \zeta_x dx$$

implies that

$$\zeta^2(P) \leq \ell \int_L \zeta_x^2 dx ,$$

where ℓ is the side of the square or, integrating with respect t y,

$$\int_L \zeta^2(P) dy < \ell D(\zeta)$$

and, integrating now with respect to x,

$$\iint \zeta^2 dxdy < \ell^2 D(\zeta),$$

which is the proposition we wanted to prove.

Let R be a point on the boundary γ and consider a circle of radius ℓ about R. We can always take ℓ small enough, so that the subdomain L of G limited by the circle is a simply connected part of G.

We shall prove that

$$H_L(\zeta) \leq 4\pi^2 \ell^2 D_L(\zeta) .$$

Consider a circle C_r of radius r, $r \leq \ell$, about R.
As $\bar{X} = 0$,

$$\zeta(P) = \int \zeta_s ds \quad ,$$

where P is a point on C_r and where the integral is taken from the boundary to P along C_r.

Hence, by Schwarz's inequality

$$\zeta^2(P) \leq 2\pi \ell \int \zeta_s^2 ds \quad ,$$

where the integral is taken along the arc of C_r in L, or a fortiori, integrating both sides with respect to s,

$$\int \zeta^2(P)\ ds \leq 4\pi^2 \ell^2 \int \zeta_s^2 ds \quad ,$$

and, integrating now with respect to r,

$$H_L(\zeta) \leq 4\pi^2 \ell^2 D_L(\zeta) \quad .$$

For the functions ω_n defined above,

$$\pi h^2 [\omega_n(P) - \omega_m(P)] = \iint_c [\varphi_n(\bar{P}) - \varphi_m(\bar{P})]\ dxdy \quad ,$$

where the double integral is taken over a circle C of radius h about P. But this last equality implies that

$$\left\{\pi h^2 [\omega_n(P) - \omega_m(P)]\right\}^2 \leq \pi h^2 H_c(\varphi_n - \varphi_m) \quad .$$

But

$$H_c(\varphi_n - \varphi_m) \leq H_G(\varphi_n - \varphi_m)$$

and, as $\varphi_n - \varphi_m = 0$ on the boundary γ,

$$H_G(\varphi_n - \varphi_m) < CD(\varphi_n - \varphi_m)$$

by the lemma we have proved above.

φ_n and φ_m being admissible functions,

$$D(\varphi_n - \varphi_m) \to 0 \qquad \text{as } n, m \to \infty ,$$

hence

$H_G(\varphi_n - \varphi_m)$ and, a fortiori, $H_c(\varphi_n - \varphi_m)$ tend to 0,

which implies that

$$\omega_n(P) - \omega_m(P) \text{ also tends to } 0.$$

Hence, for a fixed h, ω_n uniformly tends to a limit function $u(P)$

We shall now prove that, for another h, we obtain the same function $u(P)$. We consider the function

$$\psi(x,y;h) = \log \frac{r}{h} + \frac{1}{2}\left(1 - \frac{r^2}{h^2}\right) \text{ for } r < h$$

$$= 0 \qquad\qquad\qquad \text{for } r > h .$$

For $r = h$ $\qquad \psi_r = 0.$

Let H be the circle of radius h about P, and K the circle of radius k. For ζ we take the function

$$\psi(h) - \psi(k) .$$

Using Green's formula,

$$D(\varphi_n, \psi) = \frac{2}{k^2} \iint_H \varphi \, dxdy .$$

And $D(\varphi_n, \zeta) = D(\varphi_n, \psi(h)) - (\varphi_n, \psi(k)) \to 0$ as $n \to \infty$.

Hence

$$\frac{1}{h^2} \iint_H \varphi_n dxdy - \frac{1}{s^2} \iint_K \varphi_n dxdy \to 0 \text{ as } n \to \infty,$$

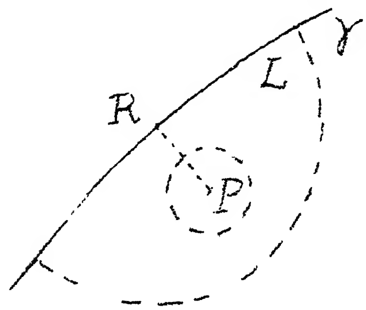
which shows that

$$u_h = u_k .$$

Hence u is defined independently of h in any open domain.

We shall show now that u has the prescribed boundary values. For ζ we take the function φ_n - which is 0 on the boundary.

Let R be the point nearest to P on the boundary, and PR=2h. The circle of radius 4h about R will determine a subdomain L in G. Let K be a circle of radius h about P. We have seen before that we can take h so small that, for all n,



$$D_L(\varphi_n - g) \leq \sigma(h),$$

where $\sigma(h)$ tends to o with h. By the lemma proved above

$$H_L(\varphi_n - g) \leq 4\pi^2(4h)^2 D_L(\varphi_n - g)$$

or

$$H_L(\;_n - g) \leq 64\pi^2 h^2 \sigma(h).$$

But

$$H_K(\varphi_n - g) \leq H_L(\varphi_n - g) .$$

Hence

$$\frac{1}{\pi h^2} H_K(\varphi_n - g) = \frac{1}{\pi h^2} \iint_K (\varphi_n - g)^2 dxdy \leq 64\pi\sigma(h) .$$

By Schwarz's inequality for integrals,

$$\left[\frac{1}{\pi h^2} \iint_K (\varphi_n - g) dxdy\right]^2 \leq \frac{1}{\pi h^2} \iint_K (\varphi_n - g)^2 dxdy \leq 64\pi\sigma(h).$$

But the left-hand side is

$$[\omega_n(P) - \breve{g}(P)]^2 ,$$

where $\breve{g}(P)$ is the mean value of g over K.

This is true for any n, hence, if $n \to \infty$, we obtain

$$[u(P) - \breve{g}(P)]^2 ,$$

and this expression tends to 0 with h. But when h tends to 0, $\tilde{g}(P)$ tends to g(P). Hence u(P) solves the boundary value condition.

We have to show now that u solves the minimum problem.

We consider a subdomain G' in G. We can cover G', except for an arbitrarily small area ε, with non-overlapping circles K_ν of radius ρ_ν, the largest radius being less than any preassigned value. If f is a function having an upper bound M in G',

$$\iint_{G'} f\,dxdy = \sum_\nu \pi\rho_\nu^2 f(P_\nu) + \delta ,$$

where P_ν is the center of the circle K_ν, and where

$$|\delta| < \varepsilon M .$$

If we take $f = u_x^2 + u^2$, then

$$D_{G'}(u) = \sum_\nu \pi\rho_\nu^2[u_x^2(P_\nu) + u_y^2(P_\nu)] + \delta$$

But

$$U_x(P_\nu) = \lim_{n\to\infty} \frac{1}{\pi\rho_\nu^2} \iint_{K_\nu} \varphi_{n_x}\,dxdy$$

and, by Schwarz's inequality,

$$u_x^2(P_\nu) \leq \frac{1}{\pi\rho_\nu^2} \iint_{K_\nu} \varphi_{n_x}^2\,dxdy .$$

Hence

$$D_{G'}(u) \leq \lim_{n\to\infty} D(\varphi_n) = d ,$$

or

$$D_G(u) = \lim_{G'\to G} D_{G'}(u) \leq d .$$

But, u being admissible,

$$D_G(u) = d.$$

Finally we shall show that u is harmonic.

$$D(u,\zeta) = 0$$

We consider the circle of radius a and the circle of radius b about P, and for ζ we take the following function:

$$\zeta = \log \frac{r}{a} \qquad \text{for} \quad b < r < a$$

$$\zeta = 0 \qquad \text{for} \quad r > a$$

and

$$\zeta = \log \frac{b}{a} \qquad \text{for} \quad r < b \quad .$$

Then

$$D(u,\zeta) = \frac{1}{a} \int_a u\,ds - \frac{1}{b} \int_b u\,ds = 0 \quad ,$$

where the integrals are taken along the circles of radius a and b. The last equality shows that u is harmonic.

Thus the second proof of Dirichlet's principle is completed.

Problems

1. Prove the triangle inequalities in the function space.

2. Prove the inequality

$$H(\zeta) \leq CD(\zeta) \quad (C > 0) \; ,$$

where ζ is assumed to be 0 only on an arc of the boundary.

In giving an existence proof for a solution to a variational problem one merely requires the existence of minimizing sequences, with suitable convergence properties. In practical applications or for purposes of computation there still remains the problem of actually constructing a minimizing sequence, and, furthermore, one which converges with a fair degree of rapidity. The method described below was first introduced by W. Ritz, who applied it to problems concerning elastic plates.*

We consider a variational integral $I(\Phi)$ defined over a function class R of admissible functions. An enumerable sequence of functions $w_1, \ldots, w_n, \ldots$ contained in the class R is said to be complete if every function Φ in R can be approximated by a finite linear combination

$$W_n = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n$$

of functions belonging to the sequence $\{w_n\}$ with preassigned accuracy. The approximation can be understood in several senses. Given any Φ in R and any ε, we want a W_n such that

a) $$|I(\Phi) - I(W_n)| < \varepsilon$$

b) $$\iint_G (\Phi - W_n)^2 \, dx \, dy < \varepsilon$$

c) $$|\Phi - W_n| < \varepsilon .$$

In the following we shall take the approximation in the sense a).

For example we know from the theory of Fourier series that the sequence of functions

---

*[illegible], Walter, "Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik," Journal für die reine und angewandte Mathematik, Vol. 135 (1908); "Theorie der [illegible]chwingungen einer quadratischen Platte mit [illegible]", Annalen der Physik, Vol. 38

$$\sin n\pi x \quad (n = 1,2,\ldots)$$

forms a complete system for all functions $\Phi(x)$ which are continuous, have a piecewise continuous derivative, and vanish at 0 and 1.

Except for the trigonometric functions, the most important and most useful complete system is given by the integral powers of x, or, in two dimensions, $x^n y^m$. The linear combinations of such functions are polynomials. Weierstrass proved the following important theorem.

If $f(x)$ is an arbitrary continuous function in a closed interval, then it may be approximated in this interval to any desired degree of accuracy by a polynomial $P_n(x)$, provided that n is taken sufficiently large.

This theorem is valid for higher dimensions as well.

Returning to the given variational integral $I(\Phi)$, we suppose, in order that the problem make sense, that the integral has a greatest lower bound d. From this follows immediately the existence of minimizing sequences Φ_n such that $I(\Phi_n) \to d$. The Ritz method consists in setting up a minimizing sequence by means of a series of auxiliary minimum problems.

We consider, for a fixed n, the integral

$$I(W_n) = I(a_1 w_1 + \ldots + a_n w_n),$$

where $\quad , w_n$ are the first n numbers of a complete system for the admissible function class R. The integral then becomes a function of the n coefficients $a_1, \ldots, a_n$ varying independently. We next consider the problem: to find the set of coefficients $a_1, \ldots a_n$ which makes $I(W_n)$ a minimum. Since I has a lower bound and depends continuously on the n parameters $a_1, \ldots, a_n$, it must attain a minimum; according to the ordinary theory of maxima and minima, the system of n equations

serves to determine the particular values $a_1 =$ which give the

minimum. We denote the minimizing function by $u_n = c_1 w_1 + \dots + c_n w_n$. The essence of the Ritz method is then contained in the following theorem:

The sequence of functions $u_1, \dots, u_n$, which are the solutions to the successive minimum problems $I(W_n)$ formed for each n, are a minimizing sequence to the original variational problem.

First, it is seen that $I(u_n)$ is a monotonically decreasing function of n, since we may regard every function W_{n-1} admissible in the (n-1)st minimum problem as an admissible function for the n-th minimum problem with the additional side condition $a_n = 0$. Therefore

$$I(u_n) \to \delta \geq d.$$

Next, the existence of a minimizing sequence $\{\Phi_n\}$ to the variational problem implies that, for some sufficiently large k,

$$I(\Phi_k) < d + \frac{\varepsilon}{2} .$$

Since the system $w_1, \dots, w_n, \dots$ is complete, there exists a suitable function $W_n = a_1 w_1 + \dots + a_n w_n$ such that

$$I(W_n) < I(\Phi_k) + \frac{\varepsilon}{2} .$$

But, by definition of u_n,

$$I(u_n) \leq I(W_n),$$

hence

$$I(u_n) < d + \varepsilon,$$

which establishes the convergence of $I(u_n)$ to d.

The process of constructing the minimizing sequence $\{u_n\}$ depends on solving the system of n equations:

$$\frac{\partial}{\partial a_i} I(W_n) = 0.$$

The process is considerably simplified if the given integral is quadratic, since in that case we have a system of linear

equations in the a's.

As an example, let us consider the case where on the boundary $\Phi = g$ is a polynomial, the boundary being given by $B(x,y) = 0$. We take for functions Φ the functions

$$\Phi = g + B(x,y)(a + bx + cy + \dots).$$

This sequence of functions Φ is a minimizing sequence, and $I(\Phi)$ is a function $Q(a,b,c,\dots)$ of the coefficients $a,b,c,\dots$, and the problem is reduced to find the minimum of Q with respect to $a,b,c,\dots$.

However, we shall now come to a method by which the functions Φ are determined directly, without going through polynomials.

Method of finite differences. The fundamental idea of this method is to replace the differential equation by a "difference equation" (an equation involving finite differences), thereby reducing the problem to a simple system of linear algebraic equations in a finite number of unknowns.

We begin by covering the xy-plane with a quadratic mesh consisting of squares of side h. To do this we draw in the plane the two sets of parallel lines

$$x = mh \qquad (m = 0,1,2,\dots)$$
$$y = nh \qquad (n = 0,1,2,\dots) .$$

These two families of lines intersect in points which we call the net, or lattice, points of the mesh.

Now, instead of considering functions of the continuous variables x and y, we consider functions which are defined only at the lattice points of the above mesh. That is, the functions are to be defined solely for the arguments $x = mh$, $y = nh$, where h is some fixed number. In any bounded domain only a finite number of lattice points will be present and hence each function will take on only a finite number of values. It is impossible to speak of the derivatives of such functions. Instead we define what we call the difference quotients of these discrete valued functions.

Let $u(x,y)$ be a function defined at the lattice points of the xy-plane. Then the forward difference quotient of u with respect to x at a lattice point (x,y) is defined to be

$$u_x(x,y) = \frac{u(x+h,y)-u(x,y)}{h}$$

and the backward difference quotient with respect to x

$$u_{\bar{x}}(x,y) = \frac{u(x,y)-u(x-h,y)}{h} .$$

In general, these two difference quotients are not equal.

In a manner similar to the above we may define the second difference quotients of a function, i.e., the difference quotients of the first difference quotients. The forward second difference quotient is given by

$$u_{\bar{x}x}(x,y) = \frac{u_{\bar{x}}(x+h,y)-u_{\bar{x}}(x,y)}{h}$$

$$= \frac{u(x+h,y)-2u(x,y)+u(x-h,y)}{h^2} .$$

Also

$$u_{x\bar{x}}(x,y) = \frac{u_x(x,y)-u_x(x-h,y)}{h}$$

$$= \frac{u(x+h,y)-2u(x,y)+u(x-h,y)}{h^2}$$

whence

$$u_{\bar{x}x} = u_{x\bar{x}} .$$

We could, if we wished, consider the second difference quotients u_{xx} and $u_{\bar{x}\bar{x}}$. However, the use of $u_{\bar{x}x}$ makes for greater symmetry.

We now replace the Laplace operator Δ by the difference operator, which we denote by Δ_h, to apply to functions defined only at the lattice points. Thus

$$\Delta_h u = u_{x\bar{x}} + u_{y\bar{y}} .$$

$$\Delta_h u = \frac{1}{h^2}[u(x+h,y) + u(x,y+h) + u(x-h,y) + u(x,y-h) - 4u(x,y)].$$

The significance of this operator becomes clear if we consider a net point P_0 and its four neighboring net points P_1, P_2, P_3, P_4. (Two net points are called neighbors if the distance between them is h.) Hence

$$\Delta_h u = \frac{1}{h^2}[u(P_1) + u(P_2) + u(P_3) + u(P_4) - 4u(P_0)].$$

That is, the value of $\Delta_h u$ at P_0 is four times the excess of the arithmetic mean of the four neighboring values over $u(P_0)$, this excess being divided by the area h^2 of the mesh.

The equation $\Delta u = 0$ corresponds to the difference equation

$$u(P_0) = \frac{u(P_1)+u(P_2)+u(P_3)+u(P_4)}{4}.$$

Hence, in a quadratic net, the Laplace equation states that the value of u at a lattice point P is the arithmetic mean of the values of u at the four neighbors of P. We have seen before that solutions of $\Delta u = 0$ possess this remarkable mean value property, where, in that case, the neighbors of P are the points of the circumference of a circle about P as a center.

Boundary value problem in a net. We cover the xy-plane with a quadratic net. Let G be any bounded domain in the xy-plane with a piecewise smooth boundary γ. The net domain G_h corresponding to the domain G consists of all the net points which lie in G. A net point is said to be a boundary point of G_h if not all of its four neighbors are in G; if all four neighboring points are in G, the point is said to be an interior point of G_h. The boundary γ_h of G_h is defined as the set of all boundary points of G_h.

To solve, with any specified degree of accuracy, the boundary value problem of the differential equation $\Delta u = 0$ for the domain G, we replace the differential equation by the difference equation $\Delta_h u = 0$ and the domain G by the corresponding net

domain G_h. If $u = g(x,y)$ is the boundary function prescribed on γ, then the boundary values at the points of γ_h are chosen as follows. If a net point P of γ_h lies on γ, then the value of $g(x,y)$ at P is taken as the value of u at P; if P does not lie on γ, we take as the value of u at P the value of $g(x,y)$ at a point of γ near to P.

We may now solve the boundary value problem for the net domain. Let N be the number of interior points of the net domain G_h. We may set up for each of the interior points the difference equation $\Delta_h u = 0$. In each case this is a linear equation involving five values of u. Some of these equations contain known quantities, i.e., those for points that are neighbors of boundary points. The other equations are homogeneous. Altogether we obtain a system of N linear equations in N unknowns, i.e., the values of u at the N interior points.

Existence and uniqueness of the solution. First we see that the maximum and minimum values of u certainly are attained on the boundary γ_h of G_h. For, if the maximum were attained at an interior point P, then the value of u at one of the four neighbors, say Q, would be at least as large as at P, because of the mean value property. If the value at Q is larger than that at P, we have a contradiction and the statement is proved. If the two values are equal, by continuing such a process about Q, we see that u must be constant in G_h, and again the statement is true. The minimum property is proved similarly.

It follows that the boundary value problems

$$\Delta_h u = 0 \quad \text{in} \quad G_h$$

$$u = 0 \quad \text{on} \quad \gamma_h$$

has the unique solution

$$u = 0 \quad \text{in} \quad G_h.$$

For this problem we have a system of N homogeneous linear equations. For the general boundary value problem

$$\Delta_h u = 0 \quad \text{in} \quad G_h$$
$$u = g \quad \text{on} \quad \gamma_h \, ,$$

we have the same N linear equations with the addition of a constant in some of them, i.e., a non-homogeneous system. From the theory of systems of linear equations the existence of the unique solution 0 for the homogeneous system implies the existence of a unique solution for the non-homogeneous system.

Practical methods. Various practical methods have been devised to solve quickly the system of N linear equations. When the domain has symmetric or special shapes, shortcuts may be found. When this is impossible general procedures can be used. These methods consist of processes of repeated manipulations which may be performed mechanically.

We begin by assuming for $u(x,y)$ at the interior net points of G_h any values whatsoever. It is desirable, however, to make this "first approximation" (which we denote by u_1) in such a way that the assumed values lie between the maximum and minimum boundary values. We now consider two procedures.

a) Order the interior net points of G_h in some arbitrary manner, $P_1, P_2, \ldots P_n$. Then replace $u_1(P_1)$ (our assumed "first approximation" at P_1) by the arithmetic mean of the assumed values u_1 at the four neighbors of P_1. Using this value, do the same for $u_1(P_2)$. Using the new values at P_1 and P_2, repeat the process for $u_1(P_3)$. Continue this process until the values of u_1 at all N of the interior points have been "corrected". We denote the corrected values by $u_2(P)$. They give a "second approximation" to the final solution. Again we start out with P_1 and proceed exactly as before, to determine a "third approximation" $u_3(P)$. This process is to be continued as long as notable differences occur between a value and its replaced value. When this is no longer the case we consider these values to be a good approximation to the actual solution of the boundary value problem.

b) Instead of proceeding as previously, where consecutive

The principle of the method may be understood from the elementary geometric concept of a vector gradient. Let $u = f(x_1,\ldots,x_n)$ be a non-negative function of the n variables x_i, or, as we might say, of the position vector $X = (x_1,\ldots,x_n)$, and let us seek to determine a vector $X = X_0$ for which u is at least stationary. We then proceed as follows: on the surface $u = f(x)$ we move a point $(x_1,\ldots,x_n,u)$ so that $x_i(t)$ and $u(t)$ become functions of a time-parameter t. Then the velocity of ascent or descent along the line $X = X(t), u = u(t)$ on the surface is

$$\frac{du}{dt} = \dot{u} = \sum_{i=1}^{n} \dot{x}_i f_{x_i} = \dot{X} \text{ grad } f .$$

We now choose the line along which the motion proceeds so that the descent is as steep as possible (lines of steepest descent).

(1) $$\dot{X} = -\text{grad } f ,$$

so that

$$\dot{u} = -(\text{grad } f)^2$$

Hence the position vector X moves according to the system of ordinary differential equations (1) along the lines of steepest descent with respect to the function f. Under very general assumptions, it is clear that X, starting from an arbitrary initial position, will, for $t \to \infty$, approach a position for which grad $f = 0$, and therefore for which f is stationary and possibly a minimum. However, instead of using the continuous procedure given by the differential equation (1), we may proceed stepwise, correcting a set of approximations x to the solutions of the equations grad $f = 0$ by corrections proportional to the respective components of -grad f.

This elementary idea can be generalized to variational problems. If we wish to determine a function $u(x,y)$ defined in G and having prescribed boundary values such that u is the solution of a variational problem

(2) $$I(v) = \iint_G F(x,y,v,v_x,v_y)dxdy = \text{min.},$$

then we interpret the desired function u as the limit for $t \to \infty$ of a function $v(x,y,t)$, whose values may be chosen arbitrarily for $t = 0$ and for all t thereafter are determined in such a way that the expression $I(v)$, considered as a function $I(t)$ of t, decreases as rapidly as possible toward its minimal value. Of course the boundary values of $v(x,y,t)$ are the same as those for $u(x,y)$, so that v_t must vanish at the boundary. If we choose $v = v(x,y,t)$, we find

(3) $$I(t) = -\iint_G v_t \; E(v)dxdy,$$

where $E(v)$ is the Euler expression corresponding to (2).

To consider a concrete example, we suppose that

$$I(v) = \iint_G (v_x^2 + v_y^2)dxdy,$$

so that our minimum problem amounts to determining the equilibrium of a membrane with given boundary deflections $g(s)$. Then $E(v) = -2\Delta v$. Incidentally (3) displays an analogy between the Euler expression and the gradient of a function $F(x_1,\dots,x_n)$ of n independent variables. The variation or "velocity" of $I(v)$ is expressed as an "inner product" of the velocity of the "independent function" v with the Euler expression $E(v)$, the gradient of a functional in function space.

We now assure ourselves of a steady descent or decrease of $I(t)$ by choosing v_t in accordance with the differential equation

(4) $$v_t = -k\,E(v),$$

where k is a positive arbitrary function of x,y. (3) then becomes

$$\dot{I}(t) = -\iint_G k[L(v)]^2 \; dxdy,$$

and again we can infer that, for $t \to \infty$, $v(x,y,t)$ will tend to the solution $u(x,y)$ of the corresponding boundary value problem $E(u) = 0$.

For the case of the membrane the differential equation (4) becomes

(5) $$v_t = \Delta v,$$

the equation of heat transfer. In our interpretation this equation describes a rapid approach to a stationary state along the "lines of steepest descent". While for the equations (4) and (5) the convergence of v for $t \to \infty$ can be proved, serious difficulty arises if we want to replace our continuous process by a process of stepwise corrections as would be required for numerical applications. Each step means a correction proportional to Δv, thus introducing higher and higher derivatives of the initial function v. Another great difficulty is presented by rigid boundary values.

Yet there do exist classes of problems where such difficulties can be overcome if the method is extended properly. First of all we may observe that it is not necessary to select the steepest descent along the gradient; it suffices to secure a safe descent at a suitably fast rate. Furthermore, if we consider problems for which the boundary value problem of the differential equations presents no difficulty for the domain G, but for which a degree of freedom in the boundary values is left, then the problem reduces to one for finding those values, and now all our difficulties disappear.

Application of the calculus of variations to the eigenvalue problems

Extremum properties of eigenvalues. Let $u_1, \dots, u_n$ be the components of a vector u, and

$$Q(u,u) = Q(u) = \sum_{\substack{i=1\\k=1}}^{n} a_{ik}u_i u_k$$

be a symmetric quadratic form, with

$$a_{ik} = a_{ki} .$$

The so-called "mixed form" is

$$Q(u,w) = \sum_{\substack{i=1\\k=1}}^{n} a_{ik}u_i w_k .$$

Let us write

$$H(u,u) = H(u) = \sum_{i=1}^{n} u_i^2 ;$$

the orthogonality condition for two vectors u and w is:

$$H(u,w) = 0 .$$

In the theory of quadratic forms it is proved that by an orthogonal transformation the vector u can be transformed into a vector v such that the form Q(u) is transformed into $\sum \lambda_i v_i^2$, with H(u) being transformed into $\sum \lambda_i v_i^2$. The λ_i are called eigenvalues or characteristic values.

These numbers can be considered as the solutions of a sequence of minimum problems.

The first problem is to make Q(u) a minimum, with the subsidiary condition H(u) = 1. The minimum, λ_1, will be attained for a certain vector u, namely e^1, and $Q(e^1) = \lambda_1$, $H(e^1) = 1$.

The second minimum problem will be to make Q(u) a minimum, with the two subsidiary conditions H(u) = 1, $H(u,e^1) = 0$. Th

minimum, λ_2, will be attained for a certain vector u, namely e^2, and $H(e^1,e^2) = 0$, $Q(e^1,e^2) = 0$, $Q(e^2) = \lambda_2$.

The k-th minimum problem will be to make $Q(u)$ a minimum, with the k subsidiary conditions $H(u) = 1$, $H(u,e^1) = 0, \ldots,$ $H(u,e^{k-1}) = 0$. The minimum, λ_k, will be attained for a certain vector u, namely e^k, and $H(e^i,e^k) = \delta_{ik}$, $\quad ,e^k) = \lambda_l$ where $\delta_{ik} = 1$ if $i = k$ and $= 0$ if $i \neq k$.

We can obtain similar results for quadratic functionals. We consider a self-adjoint partial differential equation of the second order

(1) $\quad L(u) \qquad\qquad (pu_y).$

$$(p > 0, \rho > 0)$$

where u is a function of two independent variables, x and y, defined over a domain G, of which the boundary, Γ, is a continuous curve with a piecewise continuous tangent. The boundary condition is $u = 0$ or, more generally, $\frac{\partial u}{\partial n} + \sigma u = 0$, where σ is a piecewise continuous function of the arc length on Γ and $\frac{\partial}{\partial n}$ denotes differentiation along the normal to Γ. For the variational problems equivalent to these eigenvalue problems the following quadratic functionals are to be considered:

$$E(\Phi) = D(\Phi) + \int_\Gamma p\sigma\Phi^2 ds ,$$

$$D(\Phi) = \iint_G p(\Phi_x^2 + \Phi_y^2)dxdy + \iint_G q\Phi^2 dxdy ,$$

$$H(\Phi) = \iint_G \rho\Phi^2 dxdy ,$$

and the corresponding "mixed" forms

$$E(\Phi,\Psi) = D(\Phi,\Psi) + \int_\Gamma p\sigma\Phi\Psi ds$$

$$D(\Phi,\Psi) = \iint_G p(\Phi_x\Psi_x + \Phi_y\Psi_y)dxdy + \iint_G q\Phi\Psi dxdy$$

$$H(\Phi,\Psi) = \iint_G \rho\,\Phi\Psi dxdy .$$

For these expressions we have the relations

$$E(\Phi + \Psi) = E(\Phi) + 2E(\Phi,\Psi) +$$

$$H(\Phi + \Psi) = H(\Phi) + 2H(\Phi,\Psi) +$$

A function Φ is admissible if continuous in $G + \Gamma$ and possessing piecewise continuous first derivatives.

We can obtain the eigenvalues λ_γ and the corresponding eigenfunctions u^γ of the differential equation (1) through the following minimum properties:

> Among all admissible functions the function for which the expression $E(\Phi)$ is a minimum with the subsidiary condition $H(\Phi) = 1$, is an eigenfunction u^1 of the differential equation (1) with the natural boundary condition $\frac{\partial\Phi}{\partial n} + \sigma\Phi = 0$. The minimum value of $E(\Phi)$ is the corresponding eigenvalue. If, to the condition $H(\Phi) = 1$, we add the new subsidiary condition $H(\Phi,u^1) = 0$, then the solution of this new minimum problem is an eigenfunction u^2 of (1) with the same boundary condition, and the minimum value $E(u^2) = \lambda_2$ is the corresponding eigenvalue. Generally the variational problem $E(\Phi) =$ minimum, with the subsidiary conditions $H(\Phi) = 1$ and $H(\Phi,u^i) = 0$ $(i = 1,2,\ldots k-1)$ has a solution u^k which is an eigenfunction of (1) with the boundary condition $\frac{\partial\Phi}{\partial n} + \sigma\Phi = 0$, and the minimum value $E(u^k)$ is the corresponding eigenvalue λ_n.

Instead of making $E(\Phi)$ a minimum with the condition $H(\Phi) = 1$, we can abandon this condition and make the quotient $\frac{E(\Phi)}{H(\Phi)}$ a minimum; the solution is then given with an arbitrary factor of proportionality.

We shall assume here that the minimum problems have a solution, and show that their solutions are the of partial differential equation (1).

We consider the solution u^1 of the first variational problem, $H(u^1) = 1$. Let ζ be an admissible function and ε an arbitrary constant, then $u^1 + \varepsilon\zeta$ is also an admissible function, and

$$E(u^1 + \varepsilon\zeta) \geq \lambda_1 H(u^1 + \varepsilon\zeta)$$

or

$$2\varepsilon[E(u^1,\zeta) - \lambda_1 H(u^1,\zeta)] + \varepsilon^2[E(\zeta) - \lambda_1 H(\zeta)] \geq 0$$

for every ε, which implies

$$E(u^1,\zeta) - \lambda_1 H(u^1,\zeta) = 0 .$$

Because of Green's formula

$$E(u^1,\zeta) = -\iint_G \zeta L(u^1)dxdy + \int_\Gamma p\sigma\zeta u^1 ds ,$$

and, as the function ζ is arbitrary, we obtain the equation (1) for $u = u^1$ and $\lambda = \lambda_1$.

In the second minimum problem, let η be an admissible function; then $\zeta = \eta + tu^1$ is an admissible function, and we termine t so that $H(\zeta,u^1) = 0$; we obtain $t = -H(u^1,\eta)$ Substituting in

$$E(u^2,\zeta) - \lambda_2 H(u^2,\zeta) = 0 ,$$

we obtain

$$E(u^2,\eta) - \lambda_2 H(u^2,\eta) + t[E(u^1,u^2) - \lambda_2 H(u^1,u^2)] = 0 .$$

e last term is 0, and we get

$$E(u^2,\eta) - \lambda_2 H(u^2,\eta) = 0 ,$$

ich is the same equation as in the first case. Hence u^2 eigenfunction of (1) and λ_2 the corresponding eigenvalu

In the same way, it is seen in the general case that the equation

$$E(u^i,\eta) - \lambda_i H(u^i,\eta) = 0$$

holds, η being an arbitrary admissible function. For the normal solutions of the successive minimum problems we have the relations

$$\begin{matrix} E(u^i) = \lambda_i & , & E(u^i,u^k) = 0 \\ H(u^i) = 1 & , & H(u^i,u^k) = 0 \end{matrix} \quad (i \neq k) .$$

The eigenvalues satisfy the inequality

$$\lambda_{n-1} \leq \lambda_n ,$$

for in the n-th minimum problem the domain of the functions Φ admitted to competition is not larger than in the (n-1)th problem. Hence the minimum λ_n is not smaller than the minimum λ_{n-1}.

We shall simply mention here that other eigenvalue problems can be treated with the help of the calculus of variations. The integral may be simple or multiple, and the differential equations may be of the second order or higher.

<u>The maximum-minimum property of the eigenvalues.</u> We can replace the recurrence definition of the n-th eigenvalue and the corresponding eigenfunction by a definition in which the n-th eigenvalue and the n-th eigenfunction are determined without knowing the preceding ones.

We consider the same variational problems as before, but we replace the conditions $H(\Phi,u^i) = 0$ $(i = 1,2,\ldots n-1)$ by the n-1 new conditions $H(\Phi,v^i) = 0$ $(i = 1,2,\ldots,n-1)$, where $v^1,v^2,\ldots v^{n-1}$ are any functions piecewise continuous in G. It is not decided whether the new problem has a solution. However, the expressions $D(\Phi)$ and $E(\Phi)$ have a lower bound, which depends on the functions $v^1,v^2,\ldots v^{n-1}$ and which we shall call $d(v^1,v^2,\ldots v^{n-1})$.

Given n-1 functions $v^1,v^2,\ldots v^{n-1}$, piecewise continuous in G, and $d(v^1,v^2,\ldots v^{n-1})$ being the minimum or the lower bound

of all the values that the expression $\frac{E(\Phi)}{H(\Phi)}$ can take, when Φ is any admissible function which satisfies the conditions $H(\Phi, v^i) = 0$ $(i = 1,2,\ldots,n-1)$, then λ_n is equal to the greatest value that $d(v^1, v^2, \ldots, v^{n-1})$ can take when all sets of admissible functions are taken for $v^1, v^2, \ldots v^{n-1}$. This maximum-minimum is attained for $u = u^n$ and $v^1 = u^1$, $v^2 = u^2, \ldots, v^{n-1} = u^{n-1}$.

To prove the statement, we remark first that for $v^i = u^i$ $(1 \leq i \leq n-1)$, by definition, $d(v^1, v^2, \ldots v^{n-1}) = \lambda_n$. Then, we shall show that for arbitrarily chosen $v^1, \ldots, v^{n-1}$, $d(v^1, v^2, \ldots v^{n-1}) \leq \lambda_n$. We simply have to show that there is one function Φ, satisfying the conditions $H(\Phi, v^i) = 0$ $(i = 1,2,\ldots n-1)$, for which $E(\Phi) \leq \lambda_n$. We consider a linear combination of the first n eigenvalues, $\Phi = \sum_{i=1}^{n} c_i u^i$. The n-1 relations $H(\Phi, v^i) = 0$ give n-1 linear homogeneous equations for determining the n constants $c_1, c_2, \ldots, c_n$; they can always be solved. The condition $H(\Phi) = \sum_{i=1}^{n} c_i^2 = 1$ gives the factor of proportionality. Now

$$E(\Phi) = \sum_{i,k=1}^{n} c_i c_k E(u_i, u_k) ,$$

but

$$E(u_i, u_k) = 0 \quad (i \neq k)$$

and

$$E(u_i) = \lambda_i .$$

Hence

$$E(\Phi) = \sum_{i=1}^{n} c_i^2 \lambda_i ;$$

because of $\sum_{i=1}^{n} c_i^2 = 1$ and $\lambda_n \geq \lambda_i$ $(i = 1,\ldots,n)$,

$$E(\Phi) \leq \lambda_n .$$

Hence the minimum $d(v^1, \ldots v^{n-1})$ is not greater than λ_n, and λ_n is the greatest value that the minimum can take.

The maximum-minimum property of eigenvalues is extremely useful in many physical problems. It leads immediately to two principles which we shall simply state here:

1. By strengthening the conditions in a minimum problem, the value of the minimum is not decreased, and conversely by weakening the conditions the minimum decreases or does not increase.

2. Consider two minimum problems for the same domain of admissible functions Φ. If, for each function Φ, the expression to be minimized is not smaller in the first problem than in the second, then the minimum in the first problem is not smaller than in the second.

Physically, the first principle can be stated:

When a vibratory system is forced to vibrate under certain imposed conditions, then the fundamental frequency and each harmonic can only increase. Conversely, when the conditions under which the system vibrates are weakened, the fundamental frequency and each harmonic can only decrease.

For further discussion of the extremum properties of eigenvalues we refer to Courant, R., and Hilbert, D., Methoden der Mathematischen Physik, vol. 1, chapter VI.